

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
УЧРЕЖДЕНИЕ НАУКИ

ИНСТИТУТ ФИЛОСОФИИ РОССИЙСКОЙ АКАДЕМИИ НАУК

ΕPISTEMOLOGY  
& PHILOSOPHY OF SCIENCE

**ЭПИСТЕМОЛОГИЯ И ФИЛОСОФИЯ НАУКИ**

Т. XLIV • № 2

Ежеквартальный научно-теоретический журнал

МОСКВА  
КНОРУС  
2015

---

## СОДЕРЖАНИЕ [CONTENTS]



### Editorial

- Современная аналитическая философия сознания:  
вызовы и решения [Contemporary Analytic  
Philosophy of Mind: Challenges and Solutions] . . . . . 5**

*И.Г. Гаспаров, С.М. Левин [Igor Gasparov, Sergei Levin]*



### Panel Discussion

- На пути к объяснению сознания  
[On the Way Towards Explaining Consciousness] . . . . . 20**

*Д.В. Иванов [Dmitry Ivanov]*

- Квалиа, Витгенштейн и «перевернутый спектр»  
[Qualia, Wittgenstein and "Inverted Spectrum"] . . . . . 31**

*И.Г. Гаспаров [Igor Gasparov]*

- Путь и далек, и долог... [A Long Way, Far Away...] . . . . . 36**

*И.Ф. Михайлов [Igor Michailov]*

- Феноменальные качества как реляционные свойства  
[Phenomenal Qualities as Relational Properties] . . . . . 41**

*С.М. Левин [Sergei Levin]*

- Что мы в действительности объясняем, когда пытаемся объяснить  
сознание? [What Do We Really Explain When We Try  
to Explain Consciousness?] . . . . . 47**

*М.А. Беляев [Max Belyaev]*

- Простого решения проблемы сознания не существует.  
Ответ на критику . . . . . 52**

*Д.В. Иванов [Dmitry Ivanov]*



### Epistemology and Cognition

- Experiencing the Present . . . . . 61**

*Yuri Balashov*

- О двух уровнях эпистемологии сознания  
[On the Two Levels in the Epistemology of Consciousness]. . . 74**

*Д.Н. Разеев [Danil Razeev]*

- К «прагматической» и интерактивной парадигме исследования  
интерсубъективности [Towards a Pragmatist  
and Interactive Paradigm of Studying Intersubjectivity] . . . . 87**





*Ф.А. Станжевский [Fedor Stanzhevskiy]*



### Language and Mind

- Mind and Brain States: Embedding the Mental  
in the Living Organism . . . . . 102**

*Inês Hipólito*

<b>Blockers: A Reply to Hawthorne.</b> . . . . .	112
<i>James Grindeland</i>	
<b>Почему разрыв в объяснении не является решающим аргументом против натурализма в философии сознания [Why the Explanatory Gap Is Not a Decisive Argument against Naturalism in Philosophy of Mind]</b> . . . . .	125
<i>М.А. Секацкая [Mariya Sekatskaya]</i>	
<b>Почему нас должен заботить разрыв в объяснении сознания, а не трудная проблема сознания? [Why Should We Care about the Explanatory Gap, and not the Hard Problem?]</b> . . . . .	136
<i>С.Ф. Нагуманова [Svetlana Nagumanova]</i>	
 <b>Vista</b>	
<b>Что значит быть трансценденталистом в современной аналитической философии сознания? [What Is It Like to be a Transcendentalist in the Contemporary Analytic Philosophy of Mind?]</b> . . . . .	146
<i>Д.Э. Гаспарян [Diana Gasparyan]</i>	
<b>Опровергает ли аргумент каузальных траекторий локальную супервентность ментального над физическим? [Does the Argument from Causal Trajectories Undermine Local Supervenience of Mental on Physical?]</b> . . . . .	166
<i>Д.Б. Волков [Dmitry Volkov]</i>	
 <b>Case-studies – Science studies</b>	
<b>Can We See Natural Kind Properties?</b> . . . . .	183
<i>René Jagnow</i>	
<b>Can Anosognosia Vindicate Traditionalism about Self-Deception?</b> . . . . .	206
<i>José Eduardo Porcher</i>	
 <b>Interdisciplinary Studies</b>	
<b>Phenomenal Character, Representational Content, and the Internal Correlation of Experience: Arguments Against Tracking Representationalism</b> . . . . .	218
<i>Bin Zhao</i>	
 <b>Archive</b>	
<b>Александр Бэн и его философские воззрения. Предисловие к переводу [Alexander Bain and his Philosophical Views]</b> . . . . .	230
<i>А.А. Веретенников [Andrey Veretennikov]</i>	

<b>Определение и проблемы сознания [Definition and Problems of Consciousness]</b> . . . . .	234
<i>Александр Бэн [Alexander Bain]</i>	



**Book Reviews**

<b>Современный дуализм: попытка защиты [Contemporary Dualism. An Attempt at Defense]</b> . . . . .	247
<i>И.Г. Гаспаров [Igor Gasparov]</i>	



**Наши юбиляры**

<b>Юбилей Александра Леонидовича Никифорова</b> . . . . .	252
---	-----

*Памятка для авторов* . . . . . 253

*Подписка* . . . . . 254

Публикуемые материалы прошли процедуру рецензирования и экспертного отбора.

Журнал включен в новый перечень периодических изданий, рекомендованных Высшей аттестационной комиссией РФ для публикации материалов кандидатских и докторских диссертационных исследований в области философии, социологии и культурологии (с 1 января 2007 г.).

**All materials underwent the process of anonymous peer review and were approved for publication by the Editorial Board.**

**Editor:**

Ilya Kasavin (Institute of Philosophy, Russian Academy of Sciences (IPh RAS))

**Editorial Assistants:**

Irina Gerasimova (IPh RAS)

Petr Kusliy (IPh RAS)

Liana Tukhvatulina (IPh RAS)

**Editorial Board:**

Alexandre Antonovski (IPh RAS), Vladimir Arshinov (IPh RAS), Valentin Bazhanov (Ulyanovsk State U), Irina Chernikova (Tomsk State U), Vladimir Filatov (RSUH), Steve Fuller (U of Warwick, Great Britain), Vitaly Gorokhov (IPh RAS), Vladimir Kolpakov (IPh RAS), Natalia Kuznetsova (RSUH), Jennifer Lackey (Northwestern U, USA), Joan Leach (U. of Queensland, Australia), Natalia Martishina (Siberian Transport U), Lyudmila Mikeskina (Moscow State Pedagogical U), Alexander Nikiforov (IPh RAS), Hans Poser (Technische U Berlin, Germany), Vladimir Porus (NRU Higher School of Economics), Sergei Sekundant (Odessa State U, Ukraine), Sergei Schavelev (Kursk State Medical U), Yaroslav Shramko (Kryvyi Rih National U, Ukraine)

**International Editorial Council:**

Piama Gaidenko (IPh RAS, Russia), Abdusalam Guseinov (IPh RAS, Russia), Rom Harre (London School of Economics, Great Britain), Jaakko Hintikka (Boston U, USA), Vladislav Lektorski (IPh RAS, Russia), Hans Lenk (U Karlsruhe, Germany), Vladimir Mironov (Moscow state U, Russia), Tom Rockmore (Duquesne U, USA), Vyacheslav Stepin (IPh RAS, Russia)

© Институт философии РАН. Все права защищены, 2015

© ООО «Издательство «КноРус», 2015

© Institute of Philosophy RAS. All rights reserved, 2015

© «KnoРус», 2015



Представляем специальный тематический номер нашего журнала, посвященный отдельному направлению в современной философии – аналитической философии сознания. Это первый опыт такого рода, который покажет приемлемость данной практики для будущего.

Редакторами номера и авторами редакционной статьи выступили приглашенные специалисты – И.Г. Гаспаров (Воронежская государственная медицинская академия им. Н.Н. Бурденко) и С.М. Левин (Национальный исследовательский университет Высшая школа экономики). Все материалы прошли установленную процедуру экспертного рецензирования и были рекомендованы к печати редколлегией.

## СОВРЕМЕННАЯ АНАЛИТИЧЕСКАЯ ФИЛОСОФИЯ СОЗНАНИЯ: ВЫЗОВЫ И РЕШЕНИЯ<sup>1</sup>

**Игорь Гарибович Гаспаров** – кандидат философских наук, старший преподаватель Воронежской государственной медицинской академии им. Н.Н. Бурденко. E-mail: gasparov@mail.ru **Igor**

**Gasparov** – Voronezh Medical State Academy.



В статье утверждается, что парадигма, в которой развивается аналитическая философия сознания, может быть охарактеризована тремя основными элементами: натурализм, учет данных эмпирических исследований и адекватное объяснение субъективного опыта. Популярность натурализма в современной аналитической философии сознания совмещается с его неоднородностью. В статье различаются натурализм в широком и узком смысле, а также методологический и онтологический натурализм. Учет эмпирических данных представляется точкой максимального консенсуса в современной аналитической философии сознания. Все магистральные направления философии сознания – материализм, функционализм, дуализм – включают в свою аргументацию экспериментальные данные. Объяснение субъективного опыта также многообразно – от постулирования особой онтологии ощущений до их полной элиминации. Вне зависимости от выбранной концептуальной стратегии большинство философов сегодня соглашается с тем, что без адекватного объяснения субъективного опыта у нас не может быть удовлетворительной теории сознания. В заключение статьи делается вывод о том, что сохранение текущей парадигмы зависит от успешности объяснения субъективного опыта в рамках натурализма и эмпиризма.

**Ключевые слова:** философия сознания, дуализм, материализм, физикализм, функционализм, натурализм, эмпиризм.

## CONTEMPORARY ANALYTIC PHILOSOPHY OF MIND: CHALLENGES AND SOLUTIONS

Цель настоящей статьи – описать парадигму современной аналитической философии сознания, с одной стороны, обозначив основ-

<sup>1</sup> Работа выполнена при финансовой поддержке РГНФ, проект №13-03-00427.



**Сергей Михайлович Левин** – кандидат философских наук, преподаватель департамента социологии, Национальный исследовательский университет Высшая школа экономики. E-mail: serg.m.levin@gmail.com

**Sergei Levin** – lecturer, National Research University Higher School of Economics.



In the paper, we argue that the paradigm of contemporary analytic philosophy of mind has three main theoretical elements: naturalism, respect to empirical data and the search for adequate explanation of subjective experience.

The paper begins with a brief overview of the history of the beginning of analytic philosophy of mind, with a description of its coming to Russia. Then every theoretical element of contemporary philosophy of mind is analyzed and their exemplification in the work of philosophers exposed. Naturalism is very popular in contemporary analytic philosophy of mind and it has many varieties. In the article, we specify naturalism in strong and weak senses, methodological and ontological naturalism. The shortcomings for the naturalistic in the strong sense explanation of mind are considered.

Taking into account empirical data looks like a maximum consensus in contemporary analytic philosophy of mind. All major lines in philosophy of mind – materialism, functionalism, dualism – are trying to strengthen their arguments with a help from empirical data. Materialism tend to go by neuroscience due to its theoretical foundations. Debates around functionalism today include empirical data from sociology and psychology. Moreover even the proponents of dualism, usually perceived as anti-naturalists, tend to include neuro-cognitive research in their conceptual schema.

Explaining subjective experience may take radically different forms- from postulating extra ontology of sensation to its total elimination. Nevertheless the choosing the strategy of explanation nearly everyone in the field agree that there could not be any satisfactory explanation of mind unless we have an adequate explanation of subjective experience.

In the end, we conclude that the welfare of the current paradigm of philosophy of mind depends on successful explanation of subjecting experience within the naturalistic framework and with a firm empirical data support.

**Keywords:** philosophy of mind, dualism, materialism, physicalism, functionalism, naturalism, empirism.

ные узловые точки, вокруг которых формируются главные дискуссии этой области философского дискурса, и показав многообразие позиций, присутствующих в аналитической традиции, – с другой.

С самого начала существования аналитической философии вопрос о природе сознания занимал в ней одно из центральных мест. Обычно в качестве основополагающего труда в этой области называют вышедшую в 1949 г. книгу Г. Райла «Понятие сознания» [Flanagan, and Honderich, 1995: 570–572]. Однако еще в 1921 г. вышли лекции Б. Рассела под заголовком «Анализ сознания» [Russell, 1921]. В промежутке между этими двумя публикациями можно отметить также выход целого ряда важных работ, посвященных философии сознания, среди которых следует упомянуть труды Ч. Бродя [Broad, 1925] и Г. Прайса [Price, 1932]. Тем не менее райловское «Понятие сознания» в определенном смысле можно счи-



тать поворотным моментом в формировании современной аналитической философии сознания.

В лекциях 1921 г. Рассел отмечает, что среди философов материалисты «встречались не часто» [Russell, 1921: 7]. В отношении современной философии верно как раз обратное. Материализм является наиболее распространенной позицией, тогда как на дуализм большинство смотрит с подозрением, если не с явным неприятием<sup>2</sup>. Такая ситуация едва ли была бы возможна без радикальной критики картезианства, развернутой Райлом в «Понятии сознания». Его метафора «призрака в машине» отворила дверь другой популярной метафоре, принадлежащей Д. Деннету, – «картезианский театр». Можно утверждать, что в 1950-х гг. изменяется сама парадигма, в которой развивается аналитическая философия сознания.

В Советском Союзе аналитическая философия никогда не была приоритетным направлением философских и историко-философских изысканий. То, что попадало в поле внимания исследователей, как правило, касалось логики, философии языка, теории познания, философии науки. Философия сознания оставалась по преимуществу в тени. В книгах советских философов, писавших на тему сознания ([Дубровский, 1971; Тугаринов, 1971]), можно обнаружить вопросы, схожие с теми, которыми задаются аналитические философы сознания, но это сходство требует отдельной концептуальной работы для установления тематического единства. В советской философии ссылки на ключевые фигуры аналитической философии сознания не носили сколько-нибудь массовый характер, а их труды редко включались учебные планы будущих философов-специалистов. Перевод отдельных работ (Л. Витгенштейн, Б. Рассел) не менял общего положения дел.

Сегодня ситуация с аналитической философией сознания в России меняется к лучшему. С первой половины нулевых стали регулярно проходить научные мероприятия, посвященные современной философии сознания. В МГУ, ИФ РАН, РГГУ, ВШЭ, СПбГУ появились спецкурсы по философии сознания. Переводческая активность, в которой особая роль принадлежала А.Ф. Грязнову, привела к появлению на русском языке немалого числа книг по аналитической философии сознания ([Райл, 1999; Прист, 2000; Сёрл, 2002; Деннет,

<sup>2</sup> Согласно опросу, проведенному в 2009 г., к физикалистам и им сочувствующим отнесли себя 56,5 % опрошенных, тогда как отрицают физикализм только 27,1 % [Bourge, and Chalmers, 2014].



2004; Пинкер, 2004; Чалмерс, 2013)), что позволяет как составлять достаточно представительные русскоязычные спецкурсы, так и иметь свободу выбора при освещении темы сознания в общих курсах философии. Полезным инструментом, способствующим ориентации в мире философии сознания для русскоязычного читателя, выступают словари и энциклопедии, где взгляды на ее ключевые понятия раскрываются в связи другими понятиями аналитической философии [Касавин, 2009; Никоненко, 2015]. К аналитической философии сознания в России сегодня приковано заметное исследовательское внимание. Появились монографии по философии сознания, находящиеся в русле аналитической традиции [Юлина, 2004; Васильев, 2009; 2014; Нагуманова, 2011; Волков, 2012; Иванов, 2013]. В ведущих отечественных философских журналах все чаще публикуются статьи соответствующей тематики [Иванов, 2011; Касавин, 2012; Михайлов, 2012; Мерзляков, 2013; Кузнецов, 2014; Секацкая, 2014]. Сложились устойчивые исследовательские коллективы – Московский центр исследования сознания (А.П. Беседин, В.В. Васильев, Д.В. Волков, А.В. Кузнецов и др.); Научный совет РАН по методологии искусственного интеллекта (Д.И. Дубровский, Д.В. Иванов, В.А. Лекторский и др.); Санкт-Петербургский центр философии сознания (С.М. Левин, Д.Н. Разеев, М.А. Секацкая, Д.В. Чирва и др.). При этом приходится признать, что авторы, аффилированные с российскими университетами, сегодня не относятся к разряду широко цитируемых и известных за пределами России философов сознания. Хочется надеяться, что перечисленная образовательная, переводческая и исследовательская активность рано или поздно приведет к тому, что отечественная философия сознания не просто встроится в парадигму аналитической философии сознания, но будет принимать участие в ее дальнейшем развитии.

Если попытаться сформулировать, что собой представляет эта парадигма в начале XXI в., то большинство из тех, кто сегодня занимается философией сознания, наверное, согласились бы, что объяснение сознания должно а) быть натуралистическим, б) учитывать данные эмпирических исследований и в) предлагать удовлетворительную концепцию субъективного опыта. Выполнение этих трех условий делает философию сознания весьма трудоемким и интригующим предприятием.

Почему именно эти условия? Труднее всего ответить на этот вопрос в отношении условия а. Первая половина XIX в. была временем трансцендентального идеализма à la Кант, Фихте, Шеллинг и Гегель. Вторая половина XX в. – это период





торжества натурализма: вера в науку в прошлом столетии оказалась сильнее веры во что-либо иное. Однако что такое натурализм? И являются ли все аналитические философы, пишущие о сознании, натуралистами? Ответ на второй вопрос будет отрицательным. Р. Суинберн, П. ван Инваген, Л.Р. Бейкер, Д. Зиммерман – этот список можно продолжить – будут отрицать, что они натуралисты. На первый вопрос ответить гораздо сложнее. Несмотря на то что, по словам Д. Папино, «почти каждый хочет быть натуралистом» [Papineau, 1993: 1], далеко не все понимают «натурализм» в одинаковом смысле. Г. Гассер выделяет методологический и онтологический натурализм [Gasser, 2007: 4]. Согласно методологическому натурализму, научный метод – единственный надежный источник знания. Для онтологического натуралиста существует только то, что доступно (в принципе) для научного эмпирического исследования, т.е. объекты, содержащиеся в нашем пространственно-временном континууме.

На наш взгляд, уместно также различать натурализм в широком и узком смысле. Ключевым моментом, объединяющим эти два натурализма, можно было бы считать отказ от ссылок на волю, желание или иное вмешательство сверхъестественных существ как единственное и окончательное объяснение фактического положения дел. Любое натуралистическое объяснение должно искать имманентные причины того, почему вещи именно таковы. Натурализм в узком смысле предполагает, что в качестве имманентных причин могут выступать только эмпирически устанавливаемые факты, касающиеся природы и структуры объясняемого положения дел, но не общие принципы априорного характера. Натурализм в широком смысле, напротив, будет признавать осторожное использование такого рода принципов в объяснении фактов. С точки зрения натурализма в узком смысле, например, аристотелевскую физику следовало бы квалифицировать как бесполезную спекуляцию, а с точки зрения натурализма в широком смысле – как разновидность натуралистического объяснения чувственно постигаемых явлений.

Узкое понимание натурализма накладывает серьезные ограничения на философию сознания и создает значительное напряжение, если не противоречие, между требованием натуралистического объяснения сознания и требованием предложить удовлетворительную концепцию субъективного опыта, с одной стороны, и особенностями философского метода – с другой. Т. Нагель в статье «Каково это быть летучей мышью?» вводит в философскую дискуссию понятие «пер-



спектива первого лица» и тесно связанное с этой перспективой понятие «квале» (множественное число «квалиа») [Nagel, 1974]. Если «перспектива первого лица» и «квалиа» – часть естественной реальности, то каким образом они могут быть объяснены в рамках натурализма в узком смысле? Кажется, что сама природа субъективного опыта противится такому объяснению, «перспектива первого лица» вроде бы дана нам априори до всякой эмпирической работы. Страннику натурализма в узком смысле остается только элиминировать «перспективу первого лица» и «квалиа». Однако такая стратегия не очень удачно сочетается с требованием серьезного отношения к субъективному опыту. Ведь реальность субъективного опыта представляется большинству обычных людей, так же как и многим философам, чем-то самоочевидным, чем-то, от чего невозможно отделаться просто ссылкой на ее несовместимость с научным объяснением.

Многие философы видят в концептуальном анализе существенный момент объяснения сознания. В рамках такого подхода объяснение сознания всегда будет неполным, если оно не связывает факты, установленные эмпирическими исследованиями, с нашим повседневным понятием сознания. И поскольку такой анализ будет частично априорным, натурализм в узком смысле снова оказывается в трудном положении, поскольку ему приходится объяснять, как априорная часть объяснения сознания может быть сведена к неким эмпирически устанавливаемым фактам. Здесь натурализму в узком смысле могли бы прийти на помощь идеи Крипке о том, что факты, устанавливаемые в ходе эмпирического исследования, сами являются частью значения рассматриваемого понятия. Однако, на наш взгляд, идеи Крипке не способны решить проблему в полной мере, так как априорные принципы и в его концепции метафизической необходимости апостериори продолжают играть существенную роль. Если рассмотреть его знаменитый пример с тождеством между водой и  $H_2O$ , то необходимость этого тождества немислима без априорной необходимости всякого тождества как такового. Таким образом, кажется, что принятие натурализма в узком смысле в качестве требования к объяснению сознания ведет к существенным затруднениям, и поэтому требование натуралистического объяснения предпочтительнее понимать как требование натуралистического объяснения в широком смысле.

Требование учитывать данные эмпирических исследований представляется наиболее очевидным. Хотя современные



дуалисты уделяют значительное внимание априорным аргументам (Р. Суинберн, Д. Чалмерс), большинство из них также пытается согласовать ее с результатами современных эмпирических исследований. При этом некоторые высказывают идеи о том, как могли бы выглядеть способы ее эмпирической проверки. Например, М. Бейкер и С. Гётц описывают возможный эксперимент, который помог бы нам разрешить вопрос о правоте материализма или дуализма эмпирическим путем [Baker, and Goetz, 2011: 15–16]. Так, если предположить, что гипотетическая душа может каузально воздействовать на ход событий в физическом мире, то это можно было бы проверить, проанализировав статистику нейронных событий в мозге, различия между событиями, возникающими случайно, и событиями, являющимися частью регулярной деятельности мозга. Фиксируя события, происходящие в мозге испытуемого, можно было бы найти статистическую корреляцию между активностью нейронов и произвольными и непроизвольными миганиями век человека, участвующего в эксперименте. Если бы коррелятами произвольных миганий оказалась случайная активность нейронов, это служило бы аргументом в пользу дуализма. Если же все мигания были бы скоррелированы с регулярной мозговой активностью, это было бы свидетельством в пользу материализма. Кроме того, исследователь мог бы самостоятельно стимулировать мозг испытуемого таким образом, чтобы вызывать мигание искусственно. Если бы испытуемый сообщал, что мигания произвольны, эти данные могли бы интерпретироваться в пользу гипотезы, что сознательные волевые акты являются не более чем рационализацией уже свершившихся фактов и не играют самостоятельной каузальной роли в мире, что противоречило бы исходной дуалистической гипотезе. Конечно, такой эксперимент не смог бы предоставить нам окончательное доказательство в пользу той или иной теории сознания, но вполне мог бы стать важным аргументом в споре сторон.

Кроме того, в последние 10 лет в аналитической философии сознания стали учитываться эмпирические данные нового типа, а именно: данные социологических опросов и антропологических описаний, получивших название экспериментальной философии сознания. До начала 2000-х учет данных эмпирических исследований в философии сознания подразумевал данные естественных наук, например нейрофизиологии. Предполагалось, что пролить свет на природу сознания поможет изучение мозговых процессов. Насколько же результаты такого изучения совместимы с обыденными и философскими представлениями



о сознании? Элиминативные материалисты считают, что если научные данные о работе мозга радикально отличаются от наших обыденных и философских представлений, то тем хуже для последних. Следует отказаться от того, что мы сейчас думаем о сознании так же, как мы отказались от веры в колдовство или от птолемеевой картины мира [Schouten, and de Jong, 2011: 11]. Но знаем ли мы то, от чего отказываемся?

Обыденные представления считались чем-то самоочевидным или рефлексивно доступным. Однако затем выяснилось, что в отличие от концептуальных философских вопросов вопрос о содержании обыденных представлений может иметь потенциально верифицируемый ответ. Поиском таких ответов и занялись «экспериментальные философы». Экспериментальная философия претендует на то, чтобы отнять у философов такой привычный для них аргумент, как ссылка на очевидность.

Сегодня экспериментальной философией нередко называют исследования, проливающие свет на представления обычных людей о проблемах различных областей философии – этики, эпистемологии, философии сознания и др. Одним из идейных вдохновителей экспериментальной философии сознания называют Д. Деннета. Именно он предложил для исследования сознания метод гетерофеноменологии – систематизацию описаний ментального опыта разных людей без всякой претензии на усмотрение универсалий, лежащих в основе такого опыта. Как метод гетерофеноменология противопоставляется прежде всего традиционной феноменологии, целью которой является открытие универсальных существенных характеристик сознания посредством рефлексии [Dennett, 1992: 95–98]. При отказе от субстантивации гетерофеноменологического описания оно тем не менее не может пониматься как сообщающее нам нечто существенное относительно сознания. Дело в том, что одним из центральных положений функционализма является положение о том, что обыденные интуиции относительно сознания составляют концептуальное ядро понятия сознания [Churchland, 2005]. Таким образом, систематическое изучение наших интуиций относительно сознания для некоторых функционалистов оканчивается вместе с тем и изучением сознания как такового.

К наиболее известным экспериментальным философам сознания можно причислить Д. Нобе и Д. Принца. В своей совместной работе «Интуиции о сознании: экспериментальное исследование» [Knobe, and Prinz, 2008] они утверждают, что приписывание ментальных состояний той или иной системе



зависит как от ее функциональных характеристик, так и от физических. Первые важны для того, чтобы люди могли поверить в существование у системы когнитивных состояний (вера, намерение, решение). Вторые важны для правдоподобности существования у системы феноменальных состояний (чувства счастья, боли, страха). Авторы приходят к выводу, что восприимчивость к физическим характеристикам не позволяет людям считать, что феноменальными состояниями могут обладать системы, функционально подобные людям, но радикально отличающиеся от них своим физическим устройством. В мыслительном эксперименте Н. Блока «китайская нация» скоординированные действия жителей Поднебесной коллективно эмулируют работу мозга [Block, 1978]. Результат их работы благодаря опосредующему интерфейсу выражается в поведении отдельного тела, и это поведение неотличимо от поведения обычного человека. Нобе и Принц утверждают, что люди охотно приписывают такой системе наличие у нее когнитивных аспектов сознания и не соглашаются с использованием феноменального языка описания ментальных состояний. Другими словами, люди готовы поверить, что «китайская нация» будет вести себя разумно, но не готовы согласиться с тем, что она может чувствовать боль.

Включение в философскую аргументацию методов социальных наук оказывается обоюдоострым, оно не только открывает исследователям новые пути подтверждения своих спекулятивных гипотез, но и делает уязвимыми к критике их методологические основания. Практически каждая публикация по экспериментальной философии может породить целый шлейф последующих публикаций, где для ответа на тот же вопрос используется другая методология. Например, может выясниться, что «китайская нация» Блока не обладает сознанием, поскольку этот кейс анализировался Нобе и Принцем без учета культурных и региональных особенностей. Далее, полученные результаты могут стать предметом сравнительной проверки, где будет приниматься во внимание культурный багаж респондентов. И в ходе такого сравнительного исследования может выясниться, что приписывание ощущений «китайской нации» зависит не столько от ее внутреннего устройства, сколько от места проживания того, кому мы адресуем вопросы [Huebner et al., 2010].

Сегодняшний консенсус в вопросе о роли эмпирических исследований в решении вопроса о сознании адекватно описывает Зиммерман: «Подводя итог, можно сказать, что материализм не может быть доказан просто с помощью... рассуж-



дений a priori... Дуализм находится не в лучшем положении, если принимать во внимание только доказательства a priori. Пора встать с кресла и более пристально взглянуть на содержание физического мира, чтобы понять, есть ли в нем место для нас самих» [Zimmerman, 2011: 172].

Эти слова возвращают нас к третьему элементу парадигмы современной аналитической философии сознания – необходимости предложить адекватное объяснение субъективного опыта. Трудность такого объяснения состоит в том, что, с одной стороны, субъективный опыт представляется чем-то настолько очевидным, что невозможно просто отбросить его как иллюзию. С другой стороны, он существенно отличается от научных описаний реальности. В начале 1970-х гг. Нагель [Nagel, 1974], сформулировал знаменитое определение сознания: «каково это (what it is like) быть сознающим существом, или каково это переживать определенное ментальное состояние. Как минимум с этого времени и до сих пор проблема объяснения субъективности находится в центре внимания аналитических философов сознания и вполне заслуженно получила название трудная проблемы сознания»<sup>3</sup>.

Хотя большинство современных философов соглашается с тем, что без объяснения природы субъективного опыта мы не можем иметь адекватной философии сознания, подходы к этому объяснению могут быть диаметрально противоположны. Для некоторых субъективный опыт – это не просто один из аспектов сознания, а фундаментальная основа всякого ментального состояния. Для Сёрла любая интенция имеет место лишь постольку, поскольку она может быть сознательной интенцией, т.е. попасть в поток нашего субъективного опыта [Сёрл, 2002: 69]. Субъективный опыт есть и основание, и критерий сознания; но все мыслительные операции актуально и потенциально зависят от их осознаваемости. Такой подход имеет свою метафизическую цену. Если природа сознания объясняется через субъективный опыт, последний получает особый онтологический статус (концепция дуализма свойств). Сторонники альтернативного подхода стремятся объяснить субъективный опыт исходя из факторов, к которым субъективный опыт мог бы быть редуцирован. Так, согласно интенционально-системному подходу Деннета, субъективный

<sup>3</sup> Интересно, что даже авторы, отстаивающие принципиальную нерешаемость «трудной проблемы», например «мистерианец» К. Макгинн [MacGinn, 1989], могут исходить из вполне натуралистических предпосылок. Макгинн утверждает, что эпистемологические ограничения относительно феномена сознания на человека накладывает устройство его мозга.



опыт – это не особое измерение реальности, а удобный инструмент интерпретации поведения других и самого себя [Деннет, 2004: 27–44]. Интерпретация лежит в основе объяснения Деннетом субъективного опыта и сознания вообще. В связи с этим его философия сознания отделяют от функционализма вообще и называют интерпретивизмом [Mülder, 2010]. Интерпретивизм противостоит гиперреализму относительно сознания [Dennett, 1987: 342]. По Деннету, задаваясь вопросом относительно содержания того или иного субъективного опыта, мы должны помнить, что любой ответ будет не указанием на факт субъективного опыта, а интерпретацией других фактов.

В современной литературе проблему объяснения субъективного опыта связывают в основном с двумя вопросами – о существовании нередуцируемых субъективных свойств и о субъекте этих свойств. Есть немало философов и нейрочеловек, признающих субъективность нередуцируемым аспектом реальности. В последнее время значительную известность получил «панпсихизм», согласно которому ментальные свойства являются базовыми свойствами реальности [Chalmers, 1996; Rosenberg, 2004; Koch, 2012; Tononi, 2012]. Те же, кто не готов принимать натуралистический дуализм свойств, пытаются совместить существование субъективных качеств как сущностей, не зависящих от физических характеристик реальности, но при этом и не элиминировать субъективный опыт. Несмотря на весомые аргументы против когерентности нередуктивного физикализма [Kim, 1989], Ф. Джексон и М. Тай предложили свои варианты объяснения субъективного опыта в рамках физикалистского инструментария [Jackson, 2005]. Наиболее популярной стратегией такого объяснения является истолкование феноменальных характеристик как реляционных, интенциональных свойств в противоположность «квалиа» как свойств самого сознающего субъекта. Однако остается неясным, можно ли дать адекватное материалистическое объяснение самой интенциональности.

Для тех же, кто готов признать самостоятельное существование феноменальных качеств, проблема носителей этих свойств становится важной частью повестки дня. В связи с этим понятно возрождение интереса к эмерджентизму со стороны как дуалистов, так и тех, кто симпатизирует нередуктивному материализму. Термин «эмерджентный» был введен еще во второй половине XIX в. английским философом Д.Г. Льюисом [Lewes, 1875] и широко употреблялся сторонниками «британского эмерджентизма» – С. Александером, К.Л. Морганом, Ч.Д. Броудом [McLaughlin, 1992].



В настоящее время с эмерджентизмом связаны две основные дискуссии: о природе «эмерджентных свойств» и об «эмерджентных индивидах». Само по себе существование эмерджентных свойств кажется довольно тривиальным и может быть признано даже редуccionистами. Действительно, такое свойство группы молекул воды, как, например, текучесть, является эмерджентным в том смысле, что оно невыводимо из свойств каждой отдельной из этих молекул и их взаимного расположения, но требует также учета каузальных взаимодействий между ними [Сёрл, 2002: 115–117]. Такое понимание эмерджентных свойств получило название «слабая эмерджентность». Напротив, «сильная эмерджентность» – это такое понимание эмерджентных свойств, при котором они рассматриваются как несводимые к свойствам составляющих элементов и каузальным взаимодействиям между ними и наделяются собственными каузальными силами. Более того, некоторые авторы, например У. Хаскер, полагают, что признание реальности таких свойств, как свобода воли, неизбежно влечет за собой признание существования не только сильных эмерджентных свойств, но и существования особых носителей этих свойств – «эмерджентных индивидов» [Hasker, 1999: ch. 7]. В современной литературе можно встретить несколько концепций «эмерджентных индивидов». Хаскер и Зиммерман понимают их как особые, нематериальные субстанции, являющиеся продуктом каузального взаимодействия высокоорганизованной материи (эмерджентный дуализм). Т. О'Коннор и Д. Джекобс, напротив, предлагают концепцию эмерджентных индивидов в рамках неоплатоновской метафизики, настаивая при этом на сохранении базовой материалистической онтологии [O'Connor, and Jacobs, 2003]. Несмотря на то что идея эмерджентных индивидов на первый взгляд имеет очевидные преимущества перед классическим дуализмом субстанций, ее появление объясняется подходом, господствующим в современной аналитической философии сознания. При этом сама идея особой (нематериальной) субстанции как носителя ментальных свойств, необходимым образом возникающей на основе сложных каузальных взаимодействий между материальными частицами, представляется далеко не бесспорной [Гаспаров, 2014].

Проблематика эмерджентных свойств и индивидов находится на стыке классической философии сознания и метафизики. Такое сочетание, равно как и соединение классической проблематики философии сознания с эпистемологией, также является одной из значимых характеристик современного состояния этой области философских исследований.





\*\*\*

Несмотря на огромное разнообразие тем и концепций, центральное место в современной аналитической философии сознания занимает проблема натуралистического объяснения сознания, включая натуралистическое объяснение субъективности, которое было бы способно интегрировать имеющийся на сегодняшний день массив эмпирических данных и проверенных обыденных интуиций, предоставленных как нейронаукой, так и экспериментальной психологией. От успешности решения такой задачи зависит консенсус современных аналитических философов вокруг парадигмы, которую мы попытались описать в настоящей статье.

## Библиографический список

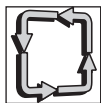
- Васильев, 2009 – *Васильев В.В.* Трудная проблема сознания. М. : Прогресс–Традиция, 2009.
- Васильев, 2014 – *Васильев В.В.* Сознание и вещи: очерк феноменалистической онтологии. М. : Либроком, 2014.
- Гаспаров, 2014 – *Гаспаров И.Г.* Проблема единства сознания и эмерджентный дуализм // Вестник ВГУ. 2014. № 11(1). С. 74–80.
- Деннет, 2004 – *Деннет Д.* Виды психики: на пути к пониманию сознания. М., 2004.
- Дубровский, 1971 – *Дубровский Д.И.* Психические явления и мозг: философский анализ проблемы в связи с актуальными задачами нейрофизиологии, психологии и кибернетики. М. : Мысль, 1971.
- Иванов, 2011 – *Иванов Д.В.* Функционализм и инверсия спектра // Эпистемология и философия науки. 2011. Т. 29, № 3. С. 82–98.
- Иванов, 2013 – *Иванов Д.В.* Природа феноменального сознания. М. : Либроком, 2013.
- Касавин, 2012 – *Касавин И.Т.* Что значит быть лондонской цветочницей? О Прометее, Пигмалионе и прочих специалистах по сознанию // Вопросы философии. 2012. № 7. С. 87–99.
- Касавин, 2009 – *Касавин И.Т.* Энциклопедия эпистемологии и философии науки. М. : Канон +, 2009.
- Кузнецов, 2014 – *Кузнецов А.В.* Когнитивные исследования и проблема ментальной каузальности // Вопросы философии. 2014. № 3. С. 133–142.
- Мерзляков, 2013 – *Мерзляков С.С.* Может ли зомби мечтать? // Эпистемология и философия науки. 2013. № 2. С. 108–122.
- Михайлов, 2012 – *Михайлов И.Ф.* Искусственный интеллект как аргумент в споре о сознании // Эпистемология и философия науки. 2012. № 2. С. 107–122.
- Нагуманова, 2011 – *Нагуманова С.Ф.* Материализм и сознание. Анализ дискуссии в современной аналитической философии. Казань : Казанский университет, 2011.



- Никоненко, 2015 – *Никоненко С.В.* Современная мультимедийная энциклопедия философских, культурологических и религиозных знаний (готовится к публикации).
- Пинкер, 2004 – *Пинкер С.* Язык как инстинкт. М. : Едиториал УРСС, 2004.
- Прист, 2000 – *Прист С.* Теории сознания. М. : Идея-Пресс : Дом интеллектуальной книги, 2000.
- Райл, 1999 – *Райл Г.* Понятие сознания. М. : Идея-Пресс, 1999.
- Секацкая, 2014 – *Секацкая М.А.* Функционализм как научная философия сознания: почему аргумент о квалиа не может быть решающим // Вопросы философии. 2014. № 3. С. 143–152.
- Сёрл, 2002 – *Сёрл Д.* Открывая сознание заново. М. : Идея-Пресс, 2002.
- Тугаринов, 1971 – *Тугаринов В.П.* Философия сознания (современные вопросы). М. : Мысль, 1971.
- Чалмерс, 2013 – *Чалмерс Д.* Сознательный ум: в поисках фундаментальной теории. М. : Editorial УРСС, 2013.
- Baker and Stewart, 2011 – *Baker M., Stewart G.* The Hypothesis of the Soul. Investigations into the Existence of the Soul. Continuum, 2011.
- Block, 1978 – *Block N.* Troubles with Functionalism // Minnesota Studies in the Philosophy of Science. 1978. № 9. P. 261–325.
- Bourge and Chalmers, 2014 – *Bourge D., Chalmers D.* What Do Philosophers Believe // Philosophical Studies. 2014. № 170. P. 465–500.
- Broad, 1925 – *Broad C.D.* The Mind and its Place in Nature. L. : Kegan Paul, 1925.
- Chalmers, 1996 – *Chalmers D.* The Conscious Mind. Oxford : Oxford University Press, 1996.
- Churchland, 2005 – *Churchland P.M.* Functionalism at Forty: A Critical Retrospective // Journal of Philosophy. 2005. № 102. P. 33–50.
- Dennett, 1987 – *Dennett D.* The Intentional Stance. Cambridge, MA : MIT Press, 1987.
- Dennett, 1992 – *Dennett D.* Consciousness Explained. N.Y. : Back Bay Books, 1992.
- Flanagan and Honderich, 1995 – The Oxford Companion to Philosophy; O. Flanagan, T. Honderich (eds.). Oxford : Oxford University Press, 1995.
- Gasser, 2007 – *Gasser G.* How Successful is Naturalism? Heusenstamm bei Frankfurt; Lancaister : Ontos Verlag, 2007.
- Hasker, 1999 – *Hasker W.* The Emergent Self. Ithaca, N.Y. : Cornell University Press, 1999.
- Huebner et al., 2010 – *Huebner B., Bruno M., Sarkissian H.* What Does the Nation of China Think about Phenomenal States? // Review of Philosophy and Psychology. 2010. № 2. P. 225–243.
- Jackson and Smith, 2005 – The Oxford Handbook of Contemporary Philosophy; F. Jackson, M. Smith (eds.). Oxford : Oxford University Press, 2005. P. 310–333.
- Jong, 2011 – *Jong H.B.* The Matter of the Mind. Malden : Blackwell Publishing, 2011.
- Kim, 1989 – *Kim J.* The Myth of Nonreductive Materialism // Proceedings and Addresses of the American Philosophical Association. 1989. № 3. P. 31–47.



- Knobe и Prinz, 2008 – *Knobe J., Prinz Js.* Intuitions about Consciousness: Experimental Studies // *Phenomenology and the Cognitive Sciences Journal of Philosophy*. 2008. № 7. P. 67–83.
- Koch, 2012 – *Koch C.* Consciousness: Confessions of a Romantical Reductionist. Cambridge : MIT Press, 2012.
- Kriegel, 2003 – *Kriegel U.* Is Intentionality Dependent Upon Consciousness? // *Philosophical Studies*. 2003. № 3. P. 271–307.
- Lewes, 1875 – *Lewes G.H.* Problems of Life and Mind. L. : Kegan Paul, Trench, Turbner, & Co, 1875.
- Lewis, 1994 – *Lewis D.* Reduction of Mind // *A Companion to Philosophy of Mind*; S. Guttenplan (ed.). Oxford : Blackwell Publishers, 1994.
- Lycan, 1987 – *Lycan W.* Consciousness. Cambridge : MIT Press, 1987.
- MacGinn, 1989 – *McGinn C.* Can We Solve the Mind-Body Problem? // *Mind*. 1989. № 391. P. 349–366.
- McLaughlin, 1992 – *McLaughlin B.P.* The Rise and Fall of British Emergentism // *Emergence or Reduction?* F.B. Beckerman, J. Kim (eds.). Berlin : Walter DeGruyter &Co, 1992.
- Mülder, 2010 – *Mülder B.* Mind Ascribed: An Elaboration and Defence of Interpretivism. Amsterdam : John Benjamins Publishing Company, 2010.
- Nagel, 1974 – *Nagel T.* What is it Like to be a Bbat? // *The Philosophical Review*. 1974. № 4. P. 435–450.
- O'Connor and Jonathan, 2003 – *O'Connor T., Jonathan J.* Emergent Individuals // *Philosophical Quarterly*. 2003. № 53. P. 540–555.
- Papineau, 1932 – *Papineau D.* Philosophical Naturalism. Oxford : Blackwell, 1993.
- Price, 1932 – *Price H.H.* Perception. Methuen and Co, 1932.
- Rosenberg, 2004 – *Rosenberg G.* A Place for Consciousness: Probing the Deep Structure of the Natural World. N.Y. : Oxford University Press, 2004.
- Russel, 1921 – *Russel B.* The Analysis of Mind. L. : George Allen and Unwin ; N.Y. : The Macmillan Company, 1921.
- Schouten and de Jong, 2011 – *Schouten M., de Jong H.B.* Mind Matters: The Roots of Reductionism // *The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience, and Reduction*; M. Schouten, H.B. de Jong (eds.). Malden, MA : Blackwell Pub., 2007.
- Tononi, 2012 – *Tononi G.* Phi: A Voyage from the Brain to the Soul. Pantheon Books, 2012.
- Zimmerman, 2011 – *Zimmerman D.* From Experience to Experiencer // *The Soul Hypothesis: Investigations into the Existence of the Soul*; M. Baker, Goetz S. (eds.). Continuum, 2011. P. 168–196.



## НА ПУТИ К ОБЪЯСНЕНИЮ СОЗНАНИЯ

**Дмитрий Валерьевич Иванов** – кандидат философских наук, старший научный сотрудник Института философии РАН. E-mail: ivdmitry@mail.ru

Статья посвящена проблеме сознания, которая заключается в выработке натуралистического объяснения феноменальных аспектов нашего сознательного опыта. Многие философы полагают, что такое объяснение невозможно. Согласно их мнению, мы не можем редуцировать феноменальные качества сознательных состояний, или квалиа, к физическим или функциональным свойствам нашего тела. Основная идея статьи состоит в следующем: для того чтобы продвинуться вперед на пути объяснения сознания, нам следует ответить на вопрос о том, являются ли квалиа свойствами, внутренне присущими сознательным состояниям, а не обсуждать, редуцируемы ли они к физическим или функциональным характеристикам. Если квалиа являются внутренне присущими свойствами, то должен быть мыслим сценарий перевернутого спектра. В статье демонстрируется, что этот сценарий невозможен. Вслед за Гилбертом Харманом в статье предлагается рассматривать квалиа как реляционные свойства репрезентируемых объектов и понимать сознательные состояния как репрезентационные состояния. Такой подход позволяет нам выработать натуралистическое объяснение сознания. В заключение отстаивается тезис о том, что эти репрезентационные состояния принадлежат не мозгу, а всему телу, вовлеченному в систему сложных взаимодействий с окружающей средой.

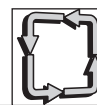
**Ключевые слова:** философия сознания, проблема сознания, феноменальное сознание, квалиа, аргумент инвертированного спектра, интенциональность, ментальные репрезентации, репрезентационизм.

## ON THE WAY TOWARDS EXPLAINING CONSCIOUSNESS

**Dmitry Ivanov** – Institute of philosophy, Russian academy of sciences.

The article deals with the problem of consciousness. It is the problem of providing naturalistic explanation of phenomenal aspects of our conscious experience. Today many philosophers believe that such explanation is impossible. According to them, we can't reduce phenomenal qualities of conscious states, or qualia, to physical or functional properties of our body. The main idea of the paper is that to move forward towards explanation of consciousness we should answer the question whether qualia are intrinsic properties of conscious states instead of discussing whether they are reducible to physical or functional characteristics. If qualia are intrinsic properties then the inverted spectrum scenario must be conceivable. In the paper it is demonstrated that this scenario is inconceivable. We can demonstrate this inconceivability appealing to Wittgenstein's critics of the theory of private language and using his arguments such as, for example, "beetle in the box" argument. If qualia are not intrinsic properties of conscious states then we must try to explain the phenomenal aspects of conscious experience using the concept of intentionality (or mental representation). Following Gilbert Harman, it is proposed to consider qualia as relational properties of represented objects and treat conscious states as representational states. This approach allows us to provide naturalistic explanation of consciousness. In conclusion it is argued that these representational states belong not to the brain alone but to the whole body that involved in complex relationship with environment.

**Key words:** philosophy of mind, the problem of consciousness, phenomenal consciousness, qualia, inverted spectrum argument, intentionality, mental representation, representationalism.



**1. Объяснение сознания должно начинаться с выделения феноменальных свойств, которыми сознательный опыт характеризуется сущностным образом.**

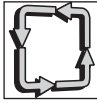
Существуют разные теории объяснения. Наиболее популярной является гемпелевская модель дедуктивно-номологического объяснения. Однако как бы ни понимать объяснение, для любой теории объяснения, как я полагаю, справедливо следующее: прежде чем объяснять природу какого-либо объекта, мы должны каким-то образом выделить этот объект, идентифицировать его как именно этот объект.

В философии ведутся споры по поводу того, каким образом осуществляется индивидуация объекта. Не вникая в метафизические дискуссии, связанные с обсуждением принципа индивидуации или закона Лейбница о тождестве неразличимых объектов, следует признать, что индивидуация, понятая как часть объяснения природы конкретного объекта, должна начинаться с фиксирования его свойств. Только специфицируя свойства конкретного объекта, мы способны начать объяснение его природы.

Все сказанное касается и проблемы объяснения сознания. Какие же свойства сознания позволяют нам выделить его как объект объяснения? Существенный вклад в выявление подобных свойств был сделан прежде всего Р. Декартом. Согласно Декарту, сознательные процессы и состояния характеризуются своей явленностью, кажимостью субъекту. Например, философ пишет: «Я – тот, кто видит свет, слышит звуки, ощущает жар. Все это – ложные ощущения, ибо я сплю. Но достоверно, что мне кажется, будто я вижу, слышу и согреваюсь. Последнее не может быть ложным, и это, собственно, то, что именуется моим ощущением» [Декарт, 1994: 25]. Иначе говоря, существенными свойствами сознательного опыта являются свойства, которые можно обозначить как феноменальные.

Если ментальные процессы и состояния характеризуются сущностным образом только наличием феноменального измерения, то их существование не может зависеть от существования каких-либо телесных объектов. Согласно Декарту, мы должны говорить о них как о модусах мыслящей вещи, особой субстанции, отличной от протяженной субстанции, которая является основой материальных объектов. Позиция Декарта, которую обозначают как дуализм субстанций, не сталкивается с проблемой объяснения сознания. Если сознание является особым супранатуралистическим объектом, то не имеет смысла искать какое-либо натуралистическое объяснение его природы. Однако против позиции дуализма субстанций было высказано уже достаточно много сильных аргументов, которые позволяют нам ее не занимать.

**2. Выделяя феноменальные свойства сознания и вместе с тем оставаясь на позициях натурализма, мы сталкиваемся с проблемой сознания.**



Отказываясь от дуализма субстанций, но сохраняя положение о том, что сознательные состояния характеризуются прежде всего феноменальными качествами, мы сталкиваемся с проблемой объяснения сознания. В самом общем виде эта проблема может быть представлена как проблема объяснения того, каким образом феноменальные качества вписываются в натуралистическую картину мира.

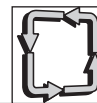
В более специфическом виде проблема представляется многими философами как проблема редукции феноменальных качеств к естественным свойствам и процессам, которые могут исследоваться средствами естественно-научных дисциплин, фундируемых физикой. Иначе говоря, объясняя сознание, мы должны показать, каким образом утверждения о феноменальных качествах дедуцируются из натуралистических утверждений. Именно вокруг проблемы редукции феноменальных свойств к физическим велись основные споры в философии сознания в последней четверти XX в. Многие философы пытались продемонстрировать нередуцируемость ментальных свойств к физическим свойствам. Позицию всех этих философов можно обозначить как дуализм свойств.

Для того чтобы разобраться, возможно ли выработать редукционистское объяснение сознания, следует прежде всего понять, что представляют собой те феноменальные качества, которые мы пытаемся объяснить. Чтобы не усложнять обсуждение данного вопроса, я не буду вдаваться в детали онтологии различных ментальных объектов. Все, что нам важно пока зафиксировать, это то, что мы как физические существа помимо материальных качеств обладаем также какими-то особыми феноменальными свойствами, которые, по мнению ряда философов, не являются физическими свойствами. Эту мысль можно сформулировать точнее. Наши сознательные состояния могут быть представлены как естественные, физические состояния. Однако, как считают дуалисты свойств, помимо своих естественных качеств эти состояния также характеризуются феноменальными свойствами, которые являются нефизическими.

Феноменальные качества, взятые в их наиболее сложном для редуктивного объяснения аспекте, часто обозначаются понятием «квалиа» (qualia)<sup>1</sup>. Это понятие призвано зафиксировать качественную составляющую нашего сознательного опыта. Скажем, когда я испытываю боль, последняя может феноменально представляться мне как режущая, колющая, ноющая и т.д.

**3. Прояснение природы феноменальных свойств заключается в решении вопроса, присущи ли внутренне (intrinsic) эти свойства нашим сознательным состояниям, а не в спорах о возможности их редукции к физическим свойствам.**

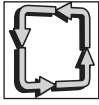
<sup>1</sup> Подробнее см.: [Иванов, 2013а; Иванов, 2013б; Туге, 2013].



Итак, для того чтобы понять сознание, нам необходимо объяснить природу феноменальных свойств, т.е. квалиа. Иначе говоря, объектом объяснения выступают свойства. Как уже отмечалось, прежде чем приступить к объяснению природы какого-либо объекта, мы должны специфицировать свойства, которые позволяют нам индивидуировать этот объект. В случае с объяснением свойств нам необходимо выявить свойства второго порядка. Описывая квалиа, многие философы помимо указания на то, что эти свойства являются феноменальной данностью, часто выделяют такие их качества, как невыразимость, приватность, некорректируемость, доступность для схватывания прямым и непосредственным образом, внутренняя присущность. Относительно того, корректно ли характеризовать феноменальные качества с помощью этих свойств, ведутся дискуссии. По моему мнению, решение этого вопроса зависит от того, принимаем ли мы положение о том, что феноменальные качества являются внутренними (intrinsic) качествами физических объектов.

Как я полагаю, вместо того чтобы решать, являются ли феноменальные качества чем-то физическим или нет, редуцируемы ли они к физическим качествам, нам следует ответить на вопрос, являются ли они внутренне присущими нам как физическим существам. Что имеется в виду под понятием «внутренняя присущность»? На этот вопрос можно ответить следующим образом. Все свойства могут быть дихотомически поделены на два класса. Некоторые свойства не принадлежат объектам самим по себе, а являются результатом того, что объекты, которым они приписываются, находятся в определенных отношениях к другим объектам. Например, какой-либо человек может обладать свойством быть мужем, только если он находится в определенных отношениях к другому человеку. Вне контекста таких отношений не имеет смысла спрашивать о том, присуще ли ему это свойство. Эти свойства могут быть обозначены как реляционные. В отличие от этого класса свойств существуют свойства, о которых мы можем говорить как о принадлежащих объектам самим по себе. Их существование не зависит от того, взаимодействует ли их носитель каким-либо образом с другими объектами. Именно эти свойства можно обозначить как внутренне присущие объектам. Понимая внутреннюю присущность подобным образом, я полагаю, что мы должны ответить отрицательно на вопрос о том, являются ли феноменальные качества свойствами, которые внутренне присущи нам как физическим существам. К такому выводу нас подводят следующие соображения.

**4. Если феноменальные свойства являются внутренними свойствами сознательных состояний, то должен быть мыслим сценарий инвертированного спектра. Однако его немыслимость можно**



**продемонстрировать с помощью рассуждений витгенштейновского типа.**

Если мы рассматриваем феноменальные качества как внутренние качества, не зависящие не только от каких-либо отношений, в которых мы как физические существа находимся к внешним объектам, но и от нашей физической организации, то должен быть мыслим сценарий инвертированного спектра. Мысленный эксперимент «инверсия спектра» в философии обсуждается достаточно долго. Уже в текстах Локка он сформулирован следующим образом.

*«Впрочем, идея голубого, имеющаяся у одного человека, может отличаться от этой идеи у другого. В наших простых идеях не было бы ничего от ложности и в том случае, если бы вследствие различного строения наших органов было бы так определено, что один и тот же предмет в одно и то же время производил бы в умах нескольких людей различные идеи; например, если бы идея, вызванная фиалкой в уме одного человека при помощи его глаз, была тождественна с идеей, вызванной в уме другого ноготками, и наоборот. Ведь этого никогда нельзя было бы узнать, потому что ум одного человека не может перейти в тело другого, чтобы воспринять, какие представления вызываются с помощью органов последнего; и потому не перепутались бы ни идеи, ни имена и ни в тех, ни в других не было бы никакой ложности. В самом деле, если все вещи, имеющие строение фиалки, будут постоянно вызывать в ком-нибудь идею, которую он назовет “голубое”, а все вещи, имеющие строение ноготков, будут постоянно вызывать идею, которую он также постоянно будет называть “желтое”, то, каковы бы ни были эти представления в его уме, он будет в состоянии так же правильно различать по ним вещи для своих надобностей и понимать и обозначать эти различия, отмеченные именами “голубое” и “желтое”, как если бы эти представления или идеи в его уме, полученные от этих двух цветков, были совершенно тождественны с идеями в умах других людей» [Локк, 1985: 444].*

В этом мысленном эксперименте нам предлагается представить ситуацию, когда два существа, тождественные физически и функционально, оказываются наделенными разными феноменальными качествами. В XX в. аргумент от инверсии спектра обсуждался прежде всего в контексте аналитической философии сознания. Несмотря на множество доводов в его поддержку, возражения, выдвинутые его оппонентами, на мой взгляд, все же заставляют нас признать, что в каком-то важном смысле перевернутый спектр немислим. Здесь нет места для подробного анализа всех дискуссий, которые возникали по поводу этого аргумента, однако общую направленность возражений можно представить следующим образом<sup>2</sup>.

<sup>2</sup> Подробнее см.: [Иванов, 2011; Иванов, 2013b].

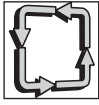




Рассуждая о статусе квалиа, мы сталкиваемся со следующей дилеммой. Если ментальные термины, обозначающие феноменальные качества, указывают на внутренние свойства сознательного опыта, то единственным критерием правильного употребления этих терминов являются свидетельства субъекта этого опыта, никто другой не может удостовериться, правильно ли фиксируются данными терминами и фиксируются ли вообще эти свойства. Однако, как отмечал Витгенштейн, критикуя идею приватного языка, в этом случае вообще не может быть каких-либо критериев правильности употребления данных ментальных терминов. Если кто-то настаивает на том, что для себя самого он правильно фиксирует этими терминами внутренние свойства своего опыта, то ему можно возразить словами Витгенштейна: «Я закрепляю для себя связь» может означать только одно: этот процесс обеспечивает то, что впоследствии я *правильно* вспоминаю эту связь. Но ведь в данном случае я не располагаю никаким критерием правильности. Так и тянет сказать: правильно то, что мне всегда представляется правильным. А это означает лишь, что здесь не может идти речь о «правильности» [Витгенштейн, 1994: 175].

Если же ментальные термины в своем употреблении ничем принципиально не отличаются от любых других слов, то онтологический статус квалиа оказывается подобен статусу жука в коробке из мысленного эксперимента Витгенштейна: «Предположим, что у каждого была бы коробка, в которой находилось бы что-то, что мы называем “жуком”. Никто не мог бы заглянуть в коробку другого; и каждый говорил бы, что он только по внешнему виду своего жука знает, что такое жук. При этом, конечно, могло бы оказаться, что в коробке у каждого находилось бы что-то другое. Можно даже представить себе, что эта вещь непрерывно изменялась бы. Ну, а если при всем том слово “жук” употреблялось бы этими людьми? В таком случае оно не было бы обозначением вещи. Вещь в коробке вообще не принадлежала бы к языковой игре даже в качестве некоего *нечто*: ведь коробка могла бы быть и пустой. Верно, тем самым вещь в этой коробке могла бы быть “сокращена”, снята независимо от того, чем бы она ни оказалась [Витгенштейн, 1994: 183].

Таким образом, феноменальные свойства, рассмотренные сторонниками возможности инверсии спектра как внутренние свойства сознательного опыта, доступ к которым имеет только субъект этого опыта, оказываются исключенными из контекстов, в которых мы способны обсуждать и сравнивать психические феномены. О них можно сказать словами Витгенштейна: «Они не нечто, но и не ничто! Вывод состоял бы лишь в том, что ничто выполняло бы такую же функцию, как и нечто, о котором ничего нельзя сказать» [Витгенштейн, 1994: 185]. Однако если мы не можем сравнивать непосредственно качества-



тивный опыт двух субъектов, то это означает, что гипотеза инвертированного спектра немыслима.

**5. Если феноменальные качества не являются внутренними свойствами нашего сознательного опыта, то феноменальные аспекты этого опыта должны мыслиться с помощью понятия интенциональности.**

Что мы имеем в виду, когда говорим, что сознательные состояния наделены феноменальными свойствами? Очевидным ответом будет указание на то, что когда мы находимся в этих состояниях, нечто явлено, дано нам определенным образом. Более того, то, что мы воспринимаем, представлено нам уникальным образом, из перспективы первого лица. Каким образом следует понимать то, что субъективные феномены воспринимаются в перспективе первого лица? Согласно мнению многих философов, это означает, что сознательный опыт интенционален. Существенной чертой интенциональных состояний является их аспектуальность. Она объясняет перспективный характер нашего опыта, т.е. тот факт, что субъективные феномены всегда даны нам определенным образом. Как пишет Сёрл, «обращать внимание на перспективный характер сознательного опыта – это хороший способ напомнить себе о том, что *всякая интенциональность аспектуальна*... Каждое интенциональное состояние имеет то, что я называю *аспектуальной формой*» [Сёрл, 2002: 131]. Сказать, что интенциональное состояние обладает аспектуальной формой, фактически означает указать на то, что это состояние имеет содержание, фиксирующее объект, на который оно направлено, в определенных аспектах.

Обсуждая проблему внутренних качеств сознательного опыта, Гилберт Харман отмечает следующие важные аспекты интенциональности [Harman, 1997]. Прежде всего наличие содержания интенционального состояния фиксирует тот факт, что субъект находится в определенном отношении к окружающей его среде. Другой важный аспект заключается в том, что содержание этих состояний не обязательно схватывает то, что действительно существует. Скажем, испанский конкистадор Понсе де Леон прибыл во Флориду в поисках фонтана вечной молодости. Он пытался найти конкретный объект, наделенный соответствующими свойствами. Об этом объекте можно говорить как об интенциональном объекте, поскольку как реальный объект он не существовал в действительности. Харман сравнивает эту ситуацию с картиной, изображающей единорога, которая, несомненно, обладает определенным содержанием. Можно сказать, что она фиксирует определенный интенциональный объект. Однако, как и фонтан вечной молодости, этот объект не существует в действительности.

Проводя параллели между изображением единорога и ментальными репрезентациями несуществующих объектов, например тех



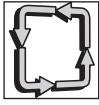
же единорогов, Харман подчеркивает, что «это не значит предположить, что ментальное воспроизведение единорога включает в себя осведомленность о свойствах самого этого ментального воспроизведения единорога». Он пишет: «Я сравниваю ментальное изображение чего-либо с картиной, но не с восприятием картины... Представление единорога отличается от представления изображения единорога... Важно проводить различие между свойствами репрезентируемого объекта и свойствами репрезентации этого объекта» [Harman, 1997: 665].

Игнорирование подобного различия приводит к представлению о существовании особых внутренних качеств сознательного опыта. Часто это представление подкрепляется аргументом от иллюзии. Харман описывает его следующим образом.

Этот аргумент начинается с непротиворечивой посылки о том, что вещи, будучи представлены в восприятии, не всегда таковы, какие они в действительности. Элоиза видит нечто коричневое и зеленое. Но перед ней нет ничего коричневого и зеленого; все это иллюзия или галлюцинация. Из этого аргумент ошибочно выводит, что коричневое и зеленое, которые видит Элоиза, не являются чем-то внешним по отношению к ней и соответственно должны быть чем-то внутренним или ментальным. Поскольку верное, не иллюзорное, не галлюцинаторное восприятие может быть качественно неотлично от иллюзорного или галлюцинаторного восприятия, в аргументе делается вывод о том, что во всех случаях восприятия Элоиза непосредственным образом осознает что-то внутреннее и ментальное и только опосредованно осознает внешние объекты, такие, как деревья и листья» [Harman, 1997: 665].

Если применить это рассуждение к картине, изображающей единорога, или к поискам фонтана вечной молодости, то следует сказать, что картина не может изображать единорога, поскольку он не существует, и что она скорее изображает что-то иное, скажем, идею единорога, имеющуюся в сознании художника. Понсе де Леон также не мог искать фонтан вечной молодости, поскольку тот не существует, скорее он искал какую-то идею этого фонтана. Однако это очевидно не так. Поиски Понсе де Леона были именно поисками определенного объекта, а не его идеи, и художник, если только он не художник-концептуалист, нацеленный на передачу на полотне идей, также стремился изобразить самого единорога.

Ошибочность данного аргумента может заключаться в необоснованном предположении о том, что восприятие в случае отсутствия в действительности объекта восприятия должно быть направлено на какой-то иной внутренний, ментальный объект. Однако если мы можем говорить о поисках, нацеленных на объект, который не существ-

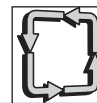


вует в действительности, то мы можем точно так же представлять и восприятие, и другие интенциональные состояния.

**6. Феноменальные качества, квалиа, которые мы считали внутренними свойствами сознательных состояний, следует рассматривать как реляционные свойства репрезентируемых объектов.**

Исходя из сказанного об интенциональности и внутренних свойствах опыта, каким образом следует мыслить феноменальные качества? Когда мы созерцаем внешний объект, скажем, дерево, то зеленый и коричневый цвета воспринимаются нами как качества этого дерева, а не как внутренние качества нашего опыта. Харман отмечает: «Посмотрите на дерево и попытайтесь направить внимание на внутренние свойства вашего визуального опыта. Вы обнаружите, что единственные свойства, на которые вы направили внимание, будут свойства дерева, включая реляционные свойства дерева, как оно видится “отсюда”» [Harman, 1997: 667]. Иначе говоря, феноменальные качества, или квалиа, являются реляционными свойствами репрезентируемых объектов. Важно подчеркнуть, что это не просто свойства объекта, а свойства объекта, наблюдаемого из определенной перспективы, или, можно сказать, свойства интенционального объекта. Именно в таком качестве они принадлежат нашему сознательному опыту.

Ситуацию восприятия объектов можно описать иным образом. Воспринимая объекты окружающей его действительности, субъект находится в определенном отношении (состоянии) к миру. Именно этот факт особого отношения к миру схватывается нами, когда мы говорим о наличии содержания у данного конкретного состояния. Содержание объясняет наличие у него перспективы, т.е. аспектуальности, которая позволяет нам говорить о данном состоянии как о феноменальном. Несомненно, мы можем говорить о факте особого отношения субъекта к миру, а значит, и о наличии определенного содержания у соответствующего репрезентационного состояния, даже когда в действительности не существует объекта, на который нацелен акт восприятия этого субъекта. Подобно книге, содержание которой представляет вымышленного героя в реальном мире, скажем, Шерлока Холмса, в содержании сознательного состояния может репрезентироваться несуществующий объект. Если вымышленный герой может обладать свойствами (например, курить трубку), то и несуществующий объект, представленный в определенном состоянии сознания, также может обладать какими-нибудь качествами. Обсуждение онтологического статуса интенциональных объектов и их свойств должно вестись, на мой взгляд, только с точки зрения того, каким содержанием обладает интенциональное состояние.

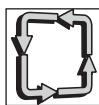


**7. Интенциональные состояния являются состояниями не мозга, а всего организма, деятельностно включенного в окружающую среду.**

Интерпретируя феноменальные качества как реляционные свойства репрезентируемого объекта, мы вполне способны выработать натуралистическое объяснение природы интенциональных состояний (ментальных репрезентаций). Как я полагаю, для этого нам нужно разобраться лишь с вопросом о том, являются ли ментальные репрезентации состояниями мозга. Последний вопрос отличается от вопроса о том, являются ли феноменальные качества внутренними свойствами ментальных репрезентаций, тем, что он требует не концептуального решения, а скорее, эмпирического. Тот, кто полагает, что ментальные состояния являются состояниями мозга, рассматривает их либо как физические, либо как функциональные состояния мозга. Однако, по мнению ряда философов и ученых, есть эмпирические свидетельства того, что онтология процессов в мозге не является онтологией ментальных состояний. Такую позицию можно обозначить как элиминативизм.

Представители элиминативизма исходят из того, что наши повседневные представления о природе психических процессов формируют определенную теорию, которую элиминативисты обозначают термином «народная психология». Подобно любой другой теории, народная психология может оказаться ложной теорией. Согласно элиминативистам, именно таковой она и является. Ее объекты, такие, как мысли, эмоции, желания и т.д., используемые для объяснения нашего когнитивного поведения, должны не объясняться посредством редукции, а элиминироваться подобно тому, как были элиминированы такие сущности, как флогистон или, скажем, *elán vital*.

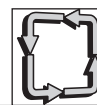
Очевидно, элиминативизм предлагает достаточно радикальный подход к пониманию психических феноменов. Такой взгляд не мог не вызвать критику. Однако, как я полагаю, не обязательно принимать все положения элиминативизма. Вполне можно допустить существование ментальных состояний, как они постулируются народной психологией, и при этом признать, что онтология процессов в мозге, обеспечивающих функционирование когнитивной сферы, не может быть понята как онтология ментальных репрезентаций, задаваемых народной психологией. Признавая справедливость этого положения элиминативизма, но не отказываясь от народной психологии, мы приходим к взгляду, согласно которому ментальные состояния, как они фиксируются народной психологией, не являются состояниями мозга. В таком случае ментальные состояния должны пониматься как состояния всего организма. Онтологически ментальные состояния являются просто положениями дел (*states of affairs*), которые заключаются в том, что организм находится в определенных отношениях с



окружающим миром. Такое понимание может быть совместимо с энактивизмом – современным подходом, разрабатываемым в контексте когнитивных наук, согласно которому сознательные состояния должны пониматься как состояния организма, зависящие не только от работы мозга, но и от его телесной организации, и от характера его деятельностной включенности в окружающую среду. В случае человека мы также можем говорить о зависимости сознательных состояний от языка и культуры.

### Библиографический список

- Витгенштейн, 1994 – *Витгенштейн Л.* Философские исследования // Л. Витгенштейн. Философские работы. М., 1994. Ч. 1. С. 75–321.
- Декарт, 1994 – *Декарт Р.* Размышления о первой философии, в коих доказывается существование Бога и различие между человеческой душой и телом // Р. Декарт. Соч. В 2 т. М., 1994. Т. 2. С. 3–73.
- Иванов, 2011 – *Иванов Д.В.* Функционализм и инверсия спектра // Эпистемология и философия науки. 2011. № 3. С. 86–98.
- Иванов, 2013а – *Иванов Д.В.* Что такое квалиа? // Сознание. Практика. Реальность: Памяти Нины Степановны Юлиной. М., 2013. С. 30–45.
- Иванов, 2013б – *Иванов Д.В.* Природа феноменального сознания. М., 2013.
- Локк, 1985 – *Локк Дж.* Опыт о человеческом разумении // Дж. Локк. Соч. В 3 т. М., 1985. Т. 1. С. 78–567.
- Сёрл, 2002 – *Сёрл Дж.* Открывая сознание заново. М., 2002.
- Harman, 1997 – *Harman G.* The Intrinsic Quality of Experience // The Nature of Consciousness. Cambridge, Mass., 1997. P. 663–677.
- Тье, 2013 – *Тье М.* Qualia // The Stanford Encyclopedia of Philosophy (Fall 2013 Edition). – <http://plato.stanford.edu/archives/fall2013/entries/qualia/>



## КВАЛИА, ВИТГЕНШТЕЙН И «ПЕРЕВЕРНУТЫЙ СПЕКТР»

**Игорь Гарибович Гаспаров** – кандидат философских наук, старший преподаватель Воронежской государственной медицинской академии им. Н.Н. Бурденко.  
E-mail: gasparov@mail.ru

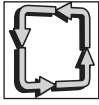
**Igor Gasparov** – Voronezh Medical State Academy.

## QUALIA, WITTGENSTEIN AND "INVERTED SPECTRUM"

В своей статье «На пути к объяснению сознания» Дмитрий Иванов приводит аргументы в пользу того, что феноменальные свойства сознания не могут быть признаны внутренними свойствами сознающего субъекта и что их следует понимать скорее как реляционные, интенциональные свойства. По его утверждению, если бы феноменальные свойства являлись внутренними свойствами, то «перевернутый спектр» был бы мыслимым сценарием. Однако некоторые рассуждения в витгенштейновском стиле позволяют нам заключить, что сценарий перевернутого спектра не имеет смысла. Следовательно, по *modus tollens*, мы должны признать, что квалиа не являются внутренними свойствами сознающего субъекта. В связи с этим возникает несколько вопросов, которые стоит обсудить подробнее.

Итак, аргумент включает две посылки. Согласно первой, мыслимость или возможность перевернутого спектра является необходимым условием существования *квалиа*, т.е. феноменальных свойств, понятых как внутренние, нереляционные свойства сознательного переживания. В дальнейшем термин «квалиа» я буду употреблять только в этом смысле, а в других случаях буду говорить просто о феноменальных свойствах. Вторая посылка утверждает, что перевернутый спектр немислим в некотором релевантном смысле этого термина. В основном я хотел бы остановиться на том, насколько хорошо вторая посылка может быть обоснована ссылкой на известные рассуждения Витгенштейна о бессмысленности приватного языка вообще и перевернутого спектра в частности. Но сначала мне кажется необходимым обратить внимание на некоторую проблематичность первой посылки аргумента Иванова.

На мой взгляд, ни немислимость, ни невозможность перевернутого спектра не могут служить решающим тестом для существования квалиа. Почему? Представьте, что любое свойство экземплифицируется объектами, имеющими это свойство, необходимым образом. В этом случае, если квалиа существуют, то они экземплифицированы необходимым образом. То есть если мое сознательное переживание экземплифицирует красный квал, то все мои двойники во всех возможных мирах тоже. Следовательно, перевернутый спектр



невозможен. Но делает ли это невозможным то, что некоторые из свойств моего феноменального переживания будут носить внутренний, нереляционный и неинтенциональный характер? *Prima facie*, кажется что нет.

Таким образом, если свойства экземплифицируются с абсолютной метафизической необходимостью, то перевернутый спектр метафизически невозможен независимо от существования квалиа. На это можно было бы возразить, что невозможность и мыслимость – не одно и то же и ключевым моментом является именно мыслимость, а не метафизическая возможность. Но я не думаю, что введение такого различия как-то спасет ситуацию, поскольку, если мы можем представить себе сценарий, в котором квалиа существуют, несмотря на отсутствие возможности перевернутого спектра, непонятно, как его мыслимость сама по себе, т.е. его представимость без соответствующей метафизической возможности, могла бы быть необходимым условием существования квалиа. Эти размышления, на мой взгляд, заставляют серьезно усомниться в оправданности первой посылки.

Обратимся ко второй посылке. Насколько витгенштейновские рассуждения способны оправдать тезис о немыслимости перевернутого спектра? Подобное обоснование проблематично по нескольким причинам. Во-первых, понятие мыслимости (*conceivability*), как оно понимается в контексте мысленных экспериментов, предполагающих перевернутый спектр, вряд ли имеет аналоги у Витгенштейна. Обычно в литературе, посвященной витгенштейновскому отношению к перевернутому спектру, указывается на то, что он отвергает его возможность исходя из верификационистских соображений. С точки зрения верификационизма некое суждение, или некий, как в данном случае, сценарий, является бессмысленным в том случае, когда для него не существует способа эмпирической проверки. В этом смысле многие, если не большинство мысленных экспериментов, обсуждаемых в современной метафизике нужно было бы признать бессмысленными. Это не удивительно, так как с точки зрения логического позитивизма сама метафизика бессмысленна.

В упрощенном варианте та же позиция приписывается Витгенштейну периода «Логико-философского трактата». Однако едва ли одной лишь ссылкой на верификационизм возможно оправдать немыслимость перевернутого спектра, тем более что в современной метафизике под мыслимостью понимается нечто совсем другое. Например, в предисловии к объемному сборнику статей под названием «*Conceivability and Possibility*» мы читаем: «У нас есть способность, которая позволяет, используя слова или концепты, или чувственные образы, репрезентировать для самих себя некие сценарии, включающие как актуальные, так и неактуальные объекты, фигурирующие как в актуальных, так и в неактуальных комбинациях» [Gendler, Haw-



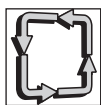


thorne, 2002: 1]. Собственно, это и есть мыслимость, и между ней и возможностью предполагается наличие некоторой и далеко не всегда однозначной связи. Понятно, что такое понимание мыслимости совершенно не совпадает с верификационизмом. Поэтому вполне уместно задать вопрос: почему необходимо опираться на верификационистский критерий при решении проблемы, обсуждаемой в существенно ином контексте? По крайней мере, здесь нужно представить обоснование для верификационистского подхода к решению проблемы мыслимости перевернутого спектра.

Кроме того, есть серьезные основания считать, что сам Витгенштейн в «Философских исследованиях», цитаты из которых приводит Иванов, отходит от верификационизма, предлагая другое понятие мыслимости и бессмысленности, основанное на характере употребления слов в языке. В § 500 он пишет: «Когда говорится, что предложение бессмысленно, то это значит, не то, что бессмыслен его смысл, а то, что определенное сочетание слов исключено из языка, из употребления» [Wittgenstein, 1993: 433]. Однако и такое понимание не соответствует общепринятой трактовке мыслимости или немыслимости определенного воображаемого сценария. Таким образом, из того, что Витгенштейн или сторонники логического позитивизма считают сценарий перевернутого спектра немыслимым, еще не следует его немыслимость в контексте метафизической дискуссии. И это не потому, что они не правы, а потому, что мыслимостью они называют разные вещи.

Другая причина, по которой, как я считаю, витгенштейновские рассуждения не способны обосновать немыслимость перевернутого спектра, носит не столько формальный, сколько содержательный характер. Дело в том, что сам философ однозначно не отрицал возможность перевернутого спектра в любой форме. В «Заметках к лекциям об “индивидуальных переживаниях” и “чувственных данных”» Витгенштейн признает, что мыслима ситуация, когда один и тот же человек по-разному воспринимает цвета. «Рассмотрим следующий случай. Кто-то говорит: “Не могу понять, сегодня все красное я вижу голубым и наоборот”. Мы отвечаем: “Это должно выглядеть странно!” Он отвечает, что это так, и продолжает разговор, например, о том, каким холодным выглядит пылающий уголь и каким раскаленным – ясное (голубое) небо. Я думаю, что при этих или сходных обстоятельствах мы склонны сказать, что он видел красным то, что мы видели голубым. Опять же мы сказали бы, что знаем, что он подразумевает под словами “голубой” и “красный” то, что подразумеваем мы, что он все время употреблял их так, как употребляли их мы» [Wittgenstein, 1968: 284].

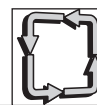
В последующей литературе сценарии, подобные описанному выше, стали называть сценариями интрасубъективного перевернутого спектра, противопоставляя их сценариям интерсубъективного пере-



вернутого спектра. Разница между ними заключается в том, что в первом случае цветовое восприятие изменяется и фиксируется внутри одного и того же субъекта, тогда как во втором случае восприятие цвета различается у двух разных субъектов. Сидни Шумейкер привел аргумент, почему если мы признаем возможность интрасубъективного перевернутого спектра, то должны признать и возможность интерсубъективного перевернутого спектра [Shoemaker, 1982: 357–381]. На мой взгляд, ничто не может воспрепятствовать этому выводу. Действительно, если предположить, что в  $t_1$  сознающий субъект  $S_1$  и отличный от него сознающий субъект  $S_2$  имеют одинаковое восприятие цвета, а в  $t_2$  восприятие цвета изменяется у  $S_1$  и остается неизменным у  $S_2$ , то имеет место не только случай интрасубъективного перевернутого спектра, но и случай интерсубъективного перевернутого спектра. Поскольку если  $S_2$  может понять вербальный отчет  $S_1$  о том, что случилось с восприятием цвета у  $S_1$ , то он также будет способен понять, что того же не случилось с его собственным восприятием цвета, и на основании этого прийти к заключению, что в  $t_2$  существуют по крайней мере два сознающих субъекта, которые по-разному воспринимают цвета одних и тех же объектов. Конечно, на основании этих рассуждений еще невозможно сделать вывод о существовании квалиа. Но на мой взгляд, их вполне достаточно, чтобы отвергнуть идею о том, что рассуждений Витгенштейна достаточно для того, чтобы утверждать немыслимость перевернутого спектра.

В заключение хотелось бы отметить и то обстоятельство, что главной целью Витгенштейна была теория «чувственных данных». Прежде всего он пытается показать, что референтами ментальных терминов типа «красный» или «зубная боль» являются некие приватные сущности, интроспективно доступные только обладающему ими субъекту. Витгенштейн пытается показать, что предположение, что такие сущности существуют, является глубоко парадоксальным и не противоречащим социальной природе языка как средства коммуникации, функционирующего только в контексте определенной деятельности. Тем не менее философ не отрицает самой данности феноменального опыта.

В тех же «Заметках к лекциям об “индивидуальных переживаниях” и “чувственных данных”» можно найти следующий пассаж: «Но что если бы я сказал, что под выражением стыда на лице я подразумевал то, что вы подразумеваете под “выражение лица + чувство”, если я явно не провожу различия между подлинными и мнимыми выражениями лица? Я считаю, что ошибочно описывать подлинное выражение как сумму выражения и чего-то еще, хотя точно такое же заблуждение – мы неправильно понимаем функцию наших выражений – возникает, если мы говорим, что подлинное выражение есть особое поведение и ничего более» [Wittgenstein, 1968: 302–303]. То есть



складывается впечатление, что Витгенштейн отрицает как адекватность аддитивного понимания феноменального переживания как некой добавки к наблюдаемому поведению, но и отрицает и отождествление этого поведения с чувством. Однако даже если мы вслед за Витгенштейном отвергаем существование чувственных данных, это еще не значит, что у нас нет возможности рассматривать феноменальные свойства как внутренне присущие качества самого субъекта. Ведь у нас еще остается возможность адвербиализма, утверждающего, что квалиа – это не что иное, как модификации самого сознающего субъекта. Несмотря на то что принятие адвербиализма, по мнению некоторых авторов, сопряжено для сторонников материализма с определенными трудностями, нельзя сказать, что эта позиция прямо ведет к субстанциальному дуализму. Поэтому кажется, что вполне возможно признать квалиа, не принимая картезианский дуализм. По крайней мере доводы Витгенштейна не ставят нас перед подобной дилеммой.

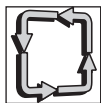
## Библиографический список

Gendler, Hawthorne, 2002 – *Gendler T.S., Hawthorne J. Conceivability and Possibility.* Oxford : Oxford University Press, 2002.

Shoemaker, 1982 – *Shoemaker S. The Inverted Spectrum // The Journal of Philosophy.* 1982. Vol. 79, No. 7.

Wittgenstein, 1968 – *Wittgenstein L. Notes for Lectures on “Private Experience” and “Sense Data” // The Philosophical Review.* 1968. Vol. 77, No. 3.

Wittgenstein, 1993 – *Wittgenstein L. Werkausgabe.* Frankfurt a/Main : Suhrkamp. Bd 1.



## ПУТЬ И ДАЛЕК, И ДОЛОГ...

**Игорь Феликсович Михайлов** – кандидат философских наук, научный сотрудник сектора методологии междисциплинарных исследований человека Института философии РАН.  
E-mail: ifmikhailov@iph.ras.ru

**Igor Michailov** – Institute of Philosophy, Russian Academy of Sciences.

Ожидание и исполнение соприкасаются в языке.

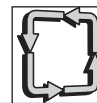
*Л. Витгенштейн.* Философские исследования

## A LONG WAY, FAR AWAY...

Читая статью Д. Иванова, я не мог отделаться от странного ощущения: я согласен почти со всеми посылками, но там, на мой взгляд, что-то не так с выводом. Главное, в чем мы согласны с автором, – это в том, что философия занимается концептуальными вопросами, оставляя наукам проблемы, для которых возможно эмпирическое решение. Но вдруг оказывается, что вся аргументация автора, выполненная на высоком профессиональном уровне, приводит к выводу, что «ментальные состояния... ментальные состояния должны пониматься как состояния всего организма». Далее я постараюсь показать, что, во-первых, этот вывод ничего не меняет в концептуальных контурах проблемы, а во-вторых, он логически не связан с выбранной автором линией аргументации.

Дмитрий исходит из того, что сознание онтологически идентифицируется через его специфические свойства, классический анализ которых восходит к Декарту. Поскольку эти свойства противоположны свойствам физических объектов или несовместимы с ними, возможны два пути решения вопроса об онтологическом статусе сознания – дуализм субстанций в духе Декарта или редукция ментальных свойств к физическим. Первый путь автор считает многократно дискредитированным и предлагает сосредоточиться на приемлемых способах редукции.

Здесь сразу возникает, на мой взгляд, законный вопрос: почему редукции именно ментального к физическому, а не наоборот? Комментируемый текст не содержит ответа, но, возможно, он лежит на поверхности: потому что физические свойства наблюдаемы, измеряемы и доступны научному исследованию, а ментальные... Но в том-то и дело, что ментальные свойства приватны, качественны и субъективны только в рамках «народной психологии», о которой весьма кстати вспоминает автор и концепцией которой мы обязаны Льюису, Деннету и др. Ненаблюдаемость феноменального опыта в достаточной степени



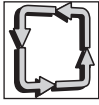
опровергается размышлениями Витгенштейна, на которые ссылается автор (приватный язык, «жук в коробке»). Современные техники психологических экспериментов позволяют количественно измерять ментальные состояния<sup>1</sup>. А бурное развитие когнитивных исследований, возможно, скоро оставит в прошлом аргумент недоступности внутренней жизни сознания для объективных наук и «других сознаний».

Думаю, вполне законно было бы предположить, что стремление части философов к физической редукции ментального вызвано не столько потребностями науки в ее нынешнем состоянии, сколько унаследованными представлениями самих философов относительно парадигмальности физического дискурса, а также их собственной – возможно, неосознаваемой – народно-психологической установкой, которая заставляет их считать, что имеет место нечто такое, что нужно редуцировать.

Далее автор указывает, что решение проблемы сознания равнозначно созданию удовлетворительной онтологии его свойств, а начать эту работу нужно с решения вопроса, являются ли феноменальные свойства внутренними или внешними, т.е. в его интерпретации, атрибутивными (присущими их носителю самому по себе) или реляционными (проявляющимися в его отношении к другим объектам). Вопрос решается в пользу реляционных свойств на основании, в частности, аргумента инвертированного спектра. Но здесь я вижу еще одну странность в рассуждениях автора. Он полагает, что если аргументы Витгенштейна верны, то «феноменальные свойства, рассмотренные сторонниками возможности инверсии спектра как внутренние свойства сознательного опыта, доступ к которым имеет только субъект этого опыта, оказываются исключенными из контекстов, в которых мы способны обсуждать и сравнивать психические феномены», а «если мы не можем сравнивать непосредственно качественный опыт двух субъектов, то это означает, что гипотеза инвертированного спектра немислима».

Во-первых, как я вижу смысл этой гипотезы, она не только вполне мыслима в контексте, но и напрямую основана на предположении о недоступности качественного опыта. Она, в том числе в формулировке Дж. Локка, которую приводит автор, говорит о возможности коммуникативного согласия между людьми относительно вторичных качеств вещей, даже если эти качества воспринимаются ими по-разному, поскольку слова выражают не квалиа, а социально сертифицированные способы различения свойств. Вокруг этой же мысли импровизирует Витгенштейн в «Философских исследованиях», в частности, рассуждая о семантике цветовых терминов (§ 239–279).

<sup>1</sup> Можно возразить, что они часто измеряются непрямым образом. А разве состояния элементарных частиц измеряются физиками непосредственно? Да и антитеза прямой–непрямой имеет смысл только в контексте «народно-психологического» убеждения в существовании скрытых и ненаблюдаемых элементов психического опыта.



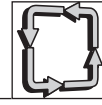
Во-вторых, Дмитрий утверждает, что если феноменальные ощущения относятся к внутренним свойствам, то они невыразимы. Но как связана природа свойств (внешние или внутренние) с их выразимостью? Мне кажется, нужно разобраться, что мы имеем в виду, когда говорим о реляционных свойствах: их материальную или формальную определенность, т.е. их проявляемость только в материальных отношениях объекта к объекту или всего лишь форму предиката (одноместный или  $n$ -местный), который выражает такое свойство (см. пример автора с «мужем»). Нетрудно заметить, что эти вещи не всегда совпадают: боль является релятивным свойством моей «души», но логически она – простое атрибутивное свойство, выражаемое одноместным предикатом. Но даже если мы директивно приведем логику болевого дискурса в соответствие со структурой фактов, т.е., например, заставим всех вместо «У меня болит голова» говорить «У меня в голове боль от спазма сосудов», это ничего не добавит к выразимости самого ощущения.

Жаль, что автор, приводя яркий пассаж Витгенштейна о том, что в популярно-психологическом дискурсе «ничто выполняло бы такую же функцию, как и нечто, о котором ничего нельзя сказать», прерывает цитирование на этих словах и оставляет вне читательского внимания следующее предложение, выражающее суть витгенштейнианской «терапии языка»: «Мы отвергаем лишь грамматику, которая здесь всячески навязывает себя нам». Большая часть «проблем сознания», таких, как квалиа, «другие сознания» и т.п., порождены естественной глубинной грамматикой наших описаний и, по мысли Витгенштейна, должны исчезнуть после правильной перенастройки языка.

Следующий шаг в рассуждениях автора таков: если неверно, что свойства сознания являются внутренними, то они интенциональны. На этом этапе я вижу следующую трудность: концепция интенциональности, созданная Францем Brentano, мультиплицировалась в многочисленные версии и варианты, не все из которых совместимы друг с другом. Помимо Гуссерля можно упомянуть Анскомб, Мейнонга, Малли, Сёрла, Деннета, кто встраивал это понятие в собственные концепции. В рассуждениях Дмитрия улавливаются явные нотки Сёрла – аспектуальность, перспектива первого лица, чего не было в оригинальной концепции Brentano. И не совсем понятно, зачем здесь эта перспектива, если многие интенциональные предикаты работают одинаково во всех лицах.

Все существующие концепции интенциональности подробно разбираются в соответствующей статье известной автору энциклопедии<sup>2</sup>, я лишь попытаюсь выразить суть проблемы в ее оригинальном

<sup>2</sup> См.: Jacob, P. Intentionality // The Stanford Encyclopedia of Philosophy; N. Zalta (ed.). – <http://plato.stanford.edu/entries/intentionality/>.



## ПУТЬ И ДАЛЕК, И ДОЛОГ...

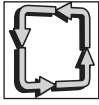
виде: как возможно, что высказывания «Я интересуюсь единорогами» и «Единорогов не существует» могут быть одновременно истинными? Brentano в свое время считал, что свойство «интенционального несуществования» присуще всем ментальным состояниям и только им (многие впоследствии оспаривали это тождество).

Как бы то ни было, остается неясным, каким образом полученный на предыдущем этапе вывод, что феноменальные свойства являются внешними реляционными свойствами сознания, имплицитно концепцию интенциональности, которая сама по себе, безусловно, уважаема и полезна. Эта неясность отчасти снимается на следующем этапе, где автор утверждает, что в контексте предыдущего логического шага феноменальные свойства оказываются свойствами не сознания, а интенциональных объектов. По крайней мере становится ясно относительно интенциональности, если уж не откуда она последовала, то для чего понадобилась. Но и здесь я не могу отделаться от вопроса: что решает это перебрасывание свойств от субъекта к объекту?

Если мы, подобно автору, исходим из того, что сознание конституируется феноменальными свойствами (что далеко не всем очевидно), то задача его научного объяснения вне картезианской скомпрометированной традиции состоит в их редукции к натуральным свойствам. Для этого на концептуально-философском уровне важно знать, свойствами чего они являются: «души», воспринимаемой ею вещи или их взаимодействия (Локк). Дмитрий говорит: они являются свойствами вещи, только интенциональной. Интенциональный объект, как он считает, — это некий «аспект» реального объекта, видимый из перспективы первого лица (ППЛ). Но тогда возникает вопрос: а из какой перспективы мы видим — если видим — неинтенциональные объекты?

Если из моей ППЛ некий натуральный объект  $O$  выглядит как интенциональный объект  $O_1$  и  $O_1$  обладает феноменальным свойством  $A$ , то присуще ли это свойство объекту  $O$ ? Если да, то введение интенциональности не означает никакого прогресса по сравнению с Локком. Если нет, то — вспомним рассуждения автора об индивидуации объектов в самом начале текста — это разные объекты. Тогда — как такое удвоение сущностей согласуется с научной картиной мира?

Но для того чтобы наконец редуцировать эти феноменально-интенциональные свойства, автор задается вопросом: являются ли ментальные репрезентации (тождественные — в его понимании — интенциональным состояниям) состояниями мозга? Здесь я хочу отметить два обстоятельства. Во-первых, слово «мозг» в этом заключительном пункте статьи появляется впервые — как гонец с важной вестью в конце пьесы. Во-вторых, сам Дмитрий признает, что этот вопрос имеет эмпирический, а не концептуальный характер и, следовательно, а priori никак не влияет на решение концептуальных проблем.

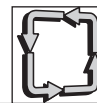


Но не это главное. В этом пункте мы должны наконец увидеть выход из концептуального лабиринта, и мы его видим. Он в том, что ментальные репрезентации не являются состояниями мозга, как мы это до сих пор, оказывается, имплицитно подразумевали. Онтологически, говорит Дмитрий, они являются определенным отношением всего организма к окружающему миру.

Итак, мозг или «весь организм»? Голова профессора Доуэля или «действительный, телесный человек, стоящий на хорошо округленной земле»? Невозможно избавиться от ощущения философского *déjà vu*, несмотря на модный «энактивизм». Распределение ментальных репрезентаций по всему организму, равно как и по всему ареалу обитания человека, ни на шаг не приближает нас к ответам на вопросы, из-за которых люди и начинают заниматься философией: действительно ли мир таков, каким я его вижу? Действительно ли другие люди думают, видят и чувствуют так же, как я? Существуют ли я, когда не мыслю? Ну и остальное – про бессмертие души и свободу воли.

Последняя фраза статьи вселяет надежду: «В случае человека мы также можем говорить о зависимости сознательных состояний от языка и культуры». Но в целом обсуждаемый текст свидетельствует, что «на пути к объяснению сознания» мы вместе с Декартом и Локком едва лишь прошли его начало.





## ФЕНОМЕНАЛЬНЫЕ КАЧЕСТВА КАК РЕЛЯЦИОННЫЕ СВОЙСТВА<sup>1</sup>

**Сергей Михайлович Левин** – кандидат философских наук, преподаватель департамента социологии, Национальный исследовательский университет Высшая школа экономики.  
E-mail: serg.m.levin@gmail.com

Непосредственное субъективное переживание определенных чувственных данных в философии сознания принято называть феноменальным качеством, или квалиа. Феноменальные качества традиционно считаются одним из наиболее трудных для натуралистического объяснения аспектов сознания. Многие философы исходят из того, что полное физическое или функциональное описание состояния человека не дает однозначного указания на то, что он в данный момент переживает на самом деле. Квалиа представляются нередуцируемыми к физическому устройству системы или к ее функциональным возможностям и зачастую объявляются особыми нефизическими свойствами. Но какими свойствами? Внутренними или реляционными? Существуют аргументы, доказывающие немислимость квалиа как внутренних свойств организма или мозга. В статье исследуется вопрос о концептуальной возможности рассмотрения квалиа как реляционных свойств организма и репрезентируемого им объекта. В заключение утверждается, что феноменальные качества как реляционные свойства при их описании могут быть сведены без потери смысла к описанию внутренних свойств организма и внутренних свойств репрезентируемого объекта, а также их натуральных некавалитативных реляционных свойств. Исходя из немислимости квалиа как свойств внутренних и редуцируемости квалиа как реляционных свойств к свойствам внутренним, делается вывод о теоретической несостоятельности самого понятия квалиа.

**Ключевые слова:** философия сознания, квалиа, внутренние свойства, реляционные свойства, ментальная репрезентация.

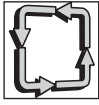
## PHENOMENAL QUALITIES AS RELATIONAL PROPERTIES

В статье Дмитрия Иванова «На пути к объяснению сознания» необходимым этапом объяснения сознания называется выделение свойств сознания, одним из которых оказывается квалитативность. В связи с этим ставится вопрос о том, являются ли феноменальные качества (квалиа) внутренними или реляционными свойствами. Иванов отстаивает точку зрения, согласно которой феноменальные качества – это реляционные свойства репрезентируемых объектов. Мы же утверждаем, что провозглашение феноменальных качеств реляционными свойствами не является чем-либо, кроме удобной языковой конвенции, которая не указывает на реально существующие, каузально эффективные реляционные свойства.

Аргументацию Дмитрия Иванова можно представить следующим образом:

- (1) Квалиа суть свойства.
- (2) Свойства бывают внутренне присущие или реляционные.

<sup>1</sup> Работа выполнена при поддержке РФФИ, проект № 13-06-00902.



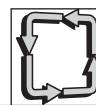
- (3) Если квалиа суть внутренние свойства сознательных состояний, то должен быть мыслим сценарий инвертированного спектра.
- (4) Сценарий инвертированного спектра немыслим.
- (5) Квалиа не суть внутренне присущие свойства.
- (6) Следовательно, квалиа суть реляционные свойства.

В статье мы сконцентрируемся на анализе положения (6) о том, что квалиа суть реляционные свойства. Противопоставление внутренних и реляционных свойств в философии имеет долгую историю. Реляционные (relational) свойства могут противопоставляться не только внутренним свойствам (intrinsic), но и внешним (extrinsic). В статье Д. Иванова реляционные и внешние свойства объединены в одну категорию. Так же поступим и мы, тем более что для целей нашей дискуссии этим различием между внешними и реляционными свойствами можно пренебречь.

Под реляционными понимаются свойства, возникающие из-за отношения одних вещей с другими. Если в отношении вовлечены две вещи, то такое свойство можно назвать диадическим, если три вещи, то триадическим, и т.д. Никогда по определению реляционное свойство не может быть свойством одной вещи, тогда это будет просто внутреннее свойство. В качестве парадигмального примера реляционного свойства в статье Дмитрия Иванова приводится свойство «быть мужем», это свойство принадлежит мужу благодаря наличию у него жены, без которой искомый человек мужем быть бы не мог. У жены очевидно нет свойства быть мужем, но у нее есть свойство быть женой, которое не принадлежит ей самой по себе и возможно, поскольку у нее есть муж.

На первый взгляд, может показаться, что есть два разных реляционных свойства – «быть мужем» для мужа и «быть женой» для жены, каждое из которых принадлежит отдельным людям, но не обоим сразу. При ближайшем рассмотрении понятно, что быть женой и быть мужем – это разные аспекты одного реляционного свойства «состоять в браке». Как отмечает Кит Файн, согласно стандартному взгляду на реляционные свойства, внутри каждого реляционного свойства можно выделить разные направления отношения этих свойств для разных участников [Fine, 2000: 2–7]. Сам Файн не согласен с тем, что направления отношения можно выделить в любом реляционном свойстве. По его мнению, есть симметричные реляционные свойства, в которых каждый участник тождествен другому в направлении отношения, но он согласен с тем, что реляционные свойства с разными направлениями отношений также существуют. В примере с реляционным свойством «состоять в браке» направлениями отношения будут «быть мужем» и «быть женой».

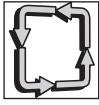
Какие же вещи Дмитрий Иванов считает участниками реляционного свойства квалиа? Одной из таких вещей, участвующей в конституи-



ровании феноменального качества, у него выступает репрезентируемый объект, о чем в статье прямо написано неоднократно. Однако, как мы помним, для экземплификации реляционного свойства одной вещи недостаточно, поэтому нельзя сказать, что феноменальные качества – это реляционные свойства только репрезентируемых объектов. Что еще помимо репрезентируемого объекта составляет исследуемое реляционные свойства? Если найдется еще одна вещь, то феноменальные качества будут диадическим реляционным свойством, если три, то триадическим, и т.д. В тексте статьи мы не нашли места, где было бы однозначно написано, что именно составляет реляционное свойство вместе с репрезентируемым объектом. Судя по общему ходу размышлений и по фразе «Онтологически ментальные состояния являются просто положениями дел (states of affairs), которые заключаются в том, что организм находится в определенных отношениях к окружающему его миру»<sup>2</sup>, вторым членом отношения можно назвать весь организм человека. Ведь все феноменальные качества, если они вообще существуют, – это ментальные состояния. Обнаружить других претендентов на участие в реляционном свойстве нам не удалось, значит, мы можем заключить, что феноменальные качества – это диадические реляционные свойства организма и репрезентируемого объекта.

С учетом наличия у реляционных свойств направлений отношений можно предположить, что феноменальное качество – это асимметричное реляционное свойство. Организм относится к репрезентируемому объекту не так, как репрезентируемый объект относится к организму. Такой объект, если мы говорим о внешней по отношению к организму неодушевленной вещи, кажется пассивным объектом, не проявляющим никакой активности для того, чтобы стать участником каких-либо отношений с организмом. Для того чтобы стать объектом феноменального ощущения, скажем, визуального опыта человека, репрезентируемый объект может обладать лишь небольшим числом необходимых и достаточных внутренних качеств (размер, прозрачность и т.п.) и его природа может очень широко варьироваться. Таким объектом может стать и небольшое насекомое, и огромная скала. Организм же, чтобы у него появилось визуальное ощущение, должен соответствовать строгому перечню необходимых требований. Можно сказать, что для появления репрезентируемого объекта в феноменальном поле организма одним из необходимых условий является деятельностная включенность последнего в окружающую среду, а со стороны репрезентируемого объекта наличие какой-либо деятельности не является необходимым условием. Направленность внимания организма на репрезентируемый объект традиционно име-

<sup>2</sup> См. с. 29 настоящего номера журнала.

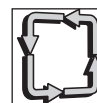


нуется интенциональностью, в обсуждаемой статье эта направленность именуется так же.

Похоже, что существуют ситуации, в которых мы можем осмысленно говорить о феноменальном качестве как о реляционном свойстве. Допустим, я вижу белую стену; в переводе с обыденного языка на язык философии сознания это означает, что ментальное состояние моего организма характеризуется наличием у него в том числе переживания квалиа белого цвета. Это феноменальное качество есть реляционное свойство моего организма и репрезентируемого объекта, а именно белой стены. Если убрать из этого отношения одну из сторон – мой организм или стену, пропадет и квалиа белого.

То, что мы можем говорить о феноменальных качествах как о реляционных свойствах, тем не менее не означает, что феноменальные качества суть реляционные свойства. Ведь иногда нам может быть просто удобно описывать некоторую ситуацию в терминах реляционных свойств, хотя эти термины будут лишь сокращенными описаниями свойств, внутренне присущих нескольким вещам, о которых мы говорим. Например, утверждение «*A* и *B* любят друг друга» звучит как описание реляционного свойства с двумя разными направлениями отношения (любить и быть любимым). По сути мы можем передать тот же смысл, апеллируя лишь к внутренним свойствам *A* и *B*. Состояние влюбленности в кого-то целиком и полностью является свойством влюбленного человека, наличие возлюбленного для этого не требуется. Дон Кихот мог воспевать свою любовь к прекрасной Дульсинее Тобосской при абсолютном отсутствии той в мире. А утверждение быть любимым, которое, кажется, относится к тому, кого любят, при ближайшем рассмотрении означает, что существует некто, любящий этого человека.

Для того чтобы мы могли сказать, что имеет место некоторое реляционное свойство, а не просто сокращенное описание нескольких внутренне присущих свойств, необходимо определить актуальную или потенциальную каузальную эффективность этого реляционного свойства, т.е. показать, что вступление в это отношение добавляет вещам что-то помимо того, что у них уже есть, или в качестве внутренних свойств, или в качестве других реляционных свойств. Отношение брака очевидно каузально значимо для супругов независимо от их внутренних свойств. Супруги, к примеру, могут претендовать на ипотечный кредит с более выгодной ставкой. А реляционное свойство пространственного взаимоположения – один лист находится над другим – также может оказывать каузальное влияние на них. Клерк возьмет для письма тот лист, что был сверху, а не снизу. Но в чем заключается каузальная эффективность феноменального качества? Как оно может повлиять на организм вне зависимости от внутренних свойств

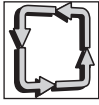


самого организма, репрезентируемого объекта и отношений между ними? Добавляет ли появление реляционного свойства квалиа что-то такое, чего уже не было в исходной ситуации?

Отношение, лежащее в основе реляционного свойства квалиа, было решено описывать при помощи термина «интенциональность». Саму интенциональность можно понимать как результат взаимодействия вещей с определенными внутренними свойствами, состоящих в соответствующих отношениях. С таким определением интенциональности может согласиться и тот, кто выделяет для нее особую субъективную онтологию (Сёрл), и тот, кто будет описывать интенциональность в объективных терминах (Деннет). Если каждый интенциональный акт может быть представлен как совокупность внутренних свойств и отношений и при этом обладать всем тем, чего мы ожидаем от феноменальных качеств без апелляции к феноменальным качествам как реляционным, то это означает, что реляционность квалиа не обладает особым онтологическим статусом.

Вернемся к примеру с белой стеной и допустим, что феноменальное переживание белого имеет место. Для появления этого квалиа, во-первых, необходимо, чтобы мой организм обладал зрением. Если бы я был незрячим, то и квалиа белого у меня бы не было. Далее можно перечислять требуемые физиологические свойства и даже включить сюда подробное описание всей нейрофизиологии. Будут ли эти описания относиться к моим внутренним качествам? Д. Льюис писал, что «предложение, или выказывание, или утверждение, которое приписывает внутренние свойства какой-либо вещи, целиком и полностью относится к этой вещи [Lewis, 1983: 197]». При таком определении внутренних свойств похоже, что утверждения, описывающие мою физиологию и нейрофизиологию, будут относиться к моим внутренним свойствам. Во-вторых, необходим более скромный список внутренних свойств стены, покрашенной в белый цвет. И, наконец, в-третьих, для того чтобы видеть стену, нам нужно находиться с ней во вполне определенных пространственных отношениях. То есть описания релевантных внутренних свойств моего организма, стены и их отношений будет достаточно, чтобы узнать о наличии у меня квалиа белого. Реляционность квалиа ничего не изменяет в онтологии ситуации.

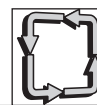
Из приведенных рассуждений напрашивается следующий вывод. Если феноменальные качества как реляционные свойства есть лишь сокращенный способ описания внутренних свойств и при этом мы соглашаемся с тем, что внутренними свойствами они тоже быть не могут, то стоит задуматься об обоснованности исследования феноменальных качеств как свойств вообще. Возможно, что реальное бытие квалиа подпадает под иную онтологическую категорию, нежели



свойства, например вещь, процесс, субстанция, или же относительно квалиа стоит занять антиреалистическую позицию и не считать, что в реальности им соответствует что-либо, кроме принятых языковых конвенций.

### Библиографический список

- Fine, 2000 – *Fine K. Neutral Relations // Philosophical Review. 2000. № 14.*  
Lewis, 1983 – *Lewis D. Extrinsic Properties // Philosophical Studies. 1983. № 44.*



## ЧТО МЫ В ДЕЙСТВИТЕЛЬНОСТИ ОБЪЯСНЯЕМ, КОГДА ПЫТАЕМСЯ ОБЪЯСНИТЬ СОЗНАНИЕ?

**Максим Александрович Беляев** – кандидат философских наук, преподаватель кафедры онтологии и теории познания Воронежского государственного университета. E-mail: [yurist84@inbox.ru](mailto:yurist84@inbox.ru)

В статье подвергнута критике позиция Д. Иванова, согласно которой натуралистическое объяснение сознания возможно осуществить посредством интерпретации феноменальных качеств как реляционных свойств репрезентируемых объектов. Показано, что не существует такого способа обоснования физического характера феноменальных качеств, чтобы данное обоснование отвечало научному идеалу объективности.

**Ключевые слова:** сознание, научное объяснение, квалиа, интенциональность, натурализм.

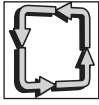
## WHAT DO WE REALLY EXPLAIN WHEN WE TRY TO EXPLAIN CONSCIOUSNESS?

**Max Belyaev** –  
Voronezh State  
University

Ivanov's aim in his article is to defend a naturalistic point of view in discussion about the nature of consciousness. The author used three philosophical presuppositions: (1) There are two kind of properties. We have some of our properties purely in virtue of the way we are (e.g., our mass). We have other properties in virtue of the way we interact with the world (e.g., our weight). The former are the intrinsic properties, the latter are the extrinsic properties; (2) The phenomenal properties are extrinsic; (3) The phenomenal properties must be ascribed to objects represented through the mind (e.g., material things, signs). The core of the article is Ivanov's argument for scientific explanation of the intentionality. This argument is strongly criticized in this paper. The point of disagreement is the following: it is impossible to find the objective explanation of consciousness since it is impossible to describe phenomenal attributes of objects as physical. The general conclusion is that if (A) to take a naturalistic and scientific explanation of consciousness as identical, and (B) to define the scientific explanation as the study the relationships between different kinds of objectively existing things, and (C) to agree that phenomenal states are relational properties of the body, there is no way to justify the physical nature of phenomenal states themselves, and therefore their objectivity. Impossibility of explaining in this case due to the fact that (1) the consciousness is not conceived at the same time objectively and subjectively existing, and (2) body do not possess intentionality.

**Key words:** consciousness, scientific explanation, qualia, intentionality, naturalism.

Д. Иванов предлагает от вопроса, редуцируемы ли феноменальные качества к физическим, перейти к вопросу, являются ли феноменальные качества внутренне присущими нам как физическим существам. Понятие внутренне присущего свойства исходит из дихотомического разделения всех свойств на



два класса – внутренние и внешние (реляционные). Внутренние свойства присущи объектам самим по себе, их существование не зависит от того, взаимодействует ли их носитель с другими объектами. Реляционные свойства, напротив, предполагают такую зависимость и не могут быть описаны вне соответствующего контекста. Рассуждая далее, Д. Иванов приходит к выводу, что феноменальные качества не являются внутренне присущими нам как физическим существам. В таком случае, согласно принятой выше классификации, следует утверждать, что феноменальные качества суть реляционные и должны быть приписаны репрезентируемому объектам (тем, что наблюдаются из определенной перспективы). В качестве таких объектов в конечном счете выступают именно объекты природного мира, вещи. Как следствие автор полагает, что «интерпретируя феноменальные качества как реляционные свойства репрезентируемого объекта, мы вполне способны выработать натуралистическое объяснение природы интенциональных состояний». С нашей точки зрения, и данное предположение, и общий ход доказательства выглядят отнюдь не бесспорными.

Отмечу принципиальный момент: с автором следует согласиться в том, что феноменальные качества для человека как физического существа не являются внутренними. Можно было бы усомниться в том, насколько качественно обоснован указанный вывод в тексте Д. Иванова, можно предложить собственное обоснование (более строгое), но важность предположений сейчас представляется второстепенной. Единственное, что хотелось бы уточнить: в каком отношении по объему находятся выражения «физическое существо» и «тело». К сожалению, автор уклонился от разъяснений на сей счет. Будем считать, что речь идет о тождестве указанных выражений, поскольку в противном случае нам пришлось бы уточнять, является ли свойство «быть физическим существом» внутренним для тела или нет. Но свойство «быть физическим существом» само может быть редуцировано к более простым свойствам, в отношении которых необходимо проделать ту же процедуру, *ad infinitum*. Такой метод представляется малопривлекательным. В то же время если в понятие «физическое существо» входит свойство сознательности (в статусе внутреннего), все рассуждения лишаются аргументативной ценности, ибо мы вновь возвращаемся к дуализму. Итак, если мы желаем, во-первых, оперировать наиболее употребительными значениями слов, а во-вторых, сохранять приверженность натуралистической парадигме, то необходимо утверждать, что физические существа суть тела. Тогда в полном согласии с выводами Д. Иванова о феноменальных качествах по отношению к телу можно утверждать одно из двух:

(А) тело обладает реляционным феноменальным свойством, и это свойство является физическим





либо

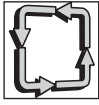
(В) тело обладает реляционным феноменальным свойством, и это свойство является нефизическим.

При этом предполагается, что все без исключения феноменальные свойства характеризуются с точки зрения данной дихотомии единым образом, т.е. являются либо физическими, либо нефизическими. Таким образом, исключена ситуация, при которой какое-либо подмножество феноменальных свойств может быть отнесено к физическим свойствам, а какое-то другое подмножество – к нефизическим.

Теперь следует уяснить, способствует ли понимание реляционного характера феноменальных свойств натуралистическому объяснению сознания. Пожалуй, предварительно стоит конкретизировать смысл, в котором используется выражение «объяснение сознания». Поскольку Д. Иванов отвергает дуалистическую концепцию сознания (во всяком случае он не видит оснований занимать данную позицию), его симпатии находятся на стороне натурализма. Как бы ни уточнять сам термин «натурализм» применительно к теории сознания, ясно одно: в этой парадигме философское объяснение сознания тождественно научному объяснению (можно было бы сказать «физическому»).

Попробуем теперь более четко определить, что является целью научного объяснения (поскольку меня интересует натуралистическая парадигма, я вправе не касаться дискуссий о различном статусе объяснения в естественных и социальных науках). Сердцевиной любого научного исследования является объяснение закономерностей, т.е. устойчивых и повторяющихся отношений между объективно существующими вещами, их свойствами и отношениями. Каким бы ни был способ объяснения, предмет объяснения остается неизменным – это объективно сущее, под которым мы понимаем то, что существует независимо от исследователя, само по себе.

Объективность – один из типов бытия сущего. От объективности следует отличать иной тип бытия, а именно субъективность. К субъективному мы относим то, с чем мы как исследователи имеем дело непосредственно (видимое мной, слышимое и т.п.). «Субъективное состояние» как понятие по содержанию тождественно феноменальному. К примеру, я своими глазами вижу кипящий чайник на газовой плите, а кроме того, знаю, что чайник со всеми его свойствами существует объективно (оставим за скобками вопрос об источниках моей уверенности). Воспринимаемое мной (чайник, издающий звук и испускающий пар) есть субъективное, но его форма, размер, содержание, а также то, что он сейчас кипит, существуют сами по себе. Очевидно, описанное положение дел не изменится, если я закрою глаза, отвернусь от плиты, выключу свет или вообще выйду из кухни. Все эти действия приведут к тому, что кипящий чайник как актуально су-



щая субъективность перестанет существовать полностью или частично, но вовсе не повлияют на то, что пламя газовой горелки нагревает воду, а она соответственно превращается в пар.

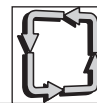
Для многих видов сущего субъективное и объективное бытие даны как бы совместно, их различие не осознается. Так, я вижу чайник и могу параллельно о нем подумать, но это все не образует никакой теоретико-познавательной сложности, поскольку в акте рефлексии я, безусловно, смогу различить два указанных способа бытия сущего. В таком случае любое научное объяснение не просто имеет дело с объективностью, но и постоянно вынуждено различать в суждении субъекта то, что касается его внутреннего опыта, и все внешнее по отношению к данному опыту. Фундаментальная способность научно объяснять нечто присутствует у субъекта в силу его способности приписывать себе феноменальные состояния и работать с ними именно как со своими состояниями.

Кажется очевидным, что натуралистическое объяснение сознания возможно лишь в ситуации (А), но не в ситуации (В), поскольку нефизический характер реляционного свойства тела вновь возвращает нас в царство дуалистической парадигмы. Но насколько возможным является объяснение феноменального состояния как физическо-го реляционного свойства тела?

Трудность состоит в следующем: от объясняющего субъекта ожидается, что он объяснит феноменальное состояние как физическое свойство, а для этого в объяснении должны различаться объективно сущее и субъективно сущее. Но я ни при каком условии не могу помыслить *мое* феноменальное состояние существующим вне меня, это абсурдно. Тогда, может быть, я должен объяснить те феноменальные состояния, которые приписываются другим? Но насколько обоснованно считать их объективными? Напомним: объективность выделяется из субъективного опыта посредством сознательной переработки этого опыта. Однако феноменальные состояния других лиц не могут выступать предметом рефлексивной деятельности сознания, так как недоступны для непосредственного восприятия. Значит, я как объясняющий субъект ничего не могу сказать о феноменальном состоянии как объективно сущем. Вероятно, я мог бы сделать феноменальные состояния предметом научного объяснения, если бы мыслил и говорил от имени тела, фактически «несубъекта», но это невозможно, поскольку несовместимо ни с сознанием, ни с речью, ни с мыслью.

Общий вывод таков: если

- (а) принимать натуралистическое объяснение сознания как тождественное научному объяснению
- и
- (б) под научным объяснением понимать изучение взаимосвязей между различными видами объективно сущего



и

(с) согласиться с тем, что феноменальные состояния являются реляционными свойствами тела (физического существа), то не существует способа обоснования физического характера самих феноменальных состояний, следовательно, и их объективности. Невозможность объяснения в данном случае связана с тем, что:

(1) сознание как вид сущего не мыслится одновременно объективно и субъективно существующим

и

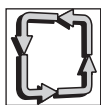
(2) тела (вещи, материальные объекты, физические существа и т.п.) не обладают интенциональностью.

Данный вывод отнюдь не исключает, что некоторые модификации приведенных выше трех базовых соглашений позволят получить иной, более позитивный результат применительно к проблематике сознания. Другими словами, моя критика не направлена ни против натурализма как определенной философской парадигмы, ни тем более против объяснения как некоторой логико-методологической процедуры. Я не исключаю, что сознание допускает такое объяснение, которое могло бы сохранять монистичность и одновременно отличаться по своему методу от научного объяснения. Однако на данный момент трудно вообразить себе подобный вариант объяснения, особенно выполняемого внутри натурализма. Если же отталкиваться от позиции Д. Иванова, то, признавая бесспорными два тезиса:

(а) феноменальные состояния суть интенциональные свойства,

(б) феноменальные состояния суть реляционные свойства,

я считаю крайне сомнительным, что эти свойства могут быть приписаны какому-либо телу вообще, будь то человеческое тело или иной предмет, взятый из окружающей нас действительности.



## П РОСТОГО РЕШЕНИЯ ПРОБЛЕМЫ СОЗНАНИЯ НЕ СУЩЕСТВУЕТ. ОТВЕТ НА КРИТИКУ

**Дмитрий Валерьевич Иванов** – кандидат философских наук, старший научный сотрудник Института философии РАН.

E-mail:  
ivdmitry@mail.ru

Статья представляет собой ответ на критику, выдвинутую в адрес работы Д.В. Иванова «На пути к объяснению сознания». Основная идея статьи заключается в том, что не существует простого решения проблемы сознания. Для того чтобы продвинуться вперед по пути выработки натуралистического объяснения сознания, нам следует оценить множество непростых аргументов и мысленных экспериментов, а также внимательно проанализировать ряд контр-интуитивных подходов к этой проблеме. Статья затрагивает такие темы, поднятые оппонентами, как необходимость и аргументы от представимости; представимость сценария инвертированного спектра; невыразимость качественных характеристик сознательных состояний; элиминация квалиа; каузальная действенность квалиа; возможность натуралистического объяснения сознания; природа интенциональности; перспективы репрезентационизма. Основной тезис, который в ней отстаивается, заключается в следующем: для того, чтобы продвинуться вперед на пути объяснения сознания, нам следует отвечать на вопрос о том, являются ли квалиа свойствами, внутренне присущими сознательным состояниям, а не пытаться редуцировать их к физическим или функциональным характеристикам организма. В статье демонстрируется, что квалиа являются не внутренними свойствами сознательных состояний, а реляционными свойствами репрезентируемых объектов. В целом, предлагаемый подход к проблеме квалиа может быть охарактеризован как комбинация элиминативизма и репрезентационизма.

**Ключевые слова:** философия сознания, проблема сознания, феноменальное сознание, квалиа, аргумент инвертированного спектра, интенциональность, ментальные репрезентации, репрезентационизм, элиминативизм.

## A SIMPLE SOLUTION TO THE PROBLEM OF CONSCIOUSNESS DOES NOT EXIST

**Dmitry Ivanov** – Institute of Philosophy, Russian Academy of Sciences

The article provides a reply to critics of *On the Way Towards Explanation of Consciousness*. The main idea of the paper is that there is no simple solution to the problem of consciousness. To move forward towards naturalistic explanation of consciousness we should examine many complicated arguments and thought experiments, carefully analyze a set of counter-intuitive approaches to this problem. The paper addresses the next issues raised by commentators: necessity and the conceivability arguments; conceivability of inverted spectrum scenario; ineffability of qualitative properties of conscious states; elimination of qualia; causal efficacy of qualia; possibility of naturalistic explanation of consciousness; the nature of intentionality; the prospects of representationalism. It defends the main thesis of *On the Way Towards Explanation of Consciousness* that to move forward towards explanation of consciousness we should answer the question whether qualia are intrinsic properties of conscious states instead of discussing whether they are reducible to physical or functional characteristics. It demonstrates that qualia are not intrinsic properties of conscious states. They are relational properties of represen-



ted objects. It is also argued that phenomenal aspects of conscious states should be explained with the concept of intentionality. The whole approach to the problem of qualia proposed in the article could be characterized as a mixture of representationalism and eliminativism.

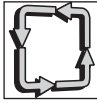
**Key words:** philosophy of mind, the problem of consciousness, phenomenal consciousness, qualia, inverted spectrum argument, intentionality, mental representation, representationalism, eliminativism

Я благодарен оппонентам за критические замечания к моей статье «На пути к объяснению сознания» (с. 20 настоящего номера). Прежде чем ответить на каждое возражение, хотелось бы сказать следующее. Я глубоко убежден, что не существует простого решения проблемы сознания, которое можно было бы представить в виде емкой формулы. Не существует также какого-то одного окончательного аргумента в поддержку подобного решения. Полагаю, объяснение природы сознательного опыта требует предварительной разработки сложного концептуального аппарата, предполагающего проведение тонких концептуальных различий, возможно, не очевидных на первый взгляд, анализ целой сети различных аргументов и мысленных экспериментов.

Такого рода работу я попытался выполнить в монографии «Природа феноменального сознания» (М., 2013). В обсуждаемой же статье, не имея возможности акцентировать внимание на употребляемых понятиях и подробно прописывать все переходы, всю аргументацию, я мог лишь кратко, тезисно обозначить основные моменты, на которые исследователям сознания следует обратить внимание. Как мне кажется, подобная краткость и породила часть критических замечаний. Надеюсь, что, отвечая на эти замечания, я смогу прояснить ключевые моменты своего подхода к проблеме сознания.

Прежде всего напомним, что в статье я дал негативный ответ на вопрос: внутренне присущи (intrinsic) ли феноменальные свойства, квалиа, нашим сознательным состояниям? Как справедливо отметил Игорь Гаспаров (с. 31), этот вывод был получен путем рассуждения, построенного по форме *modus tollens*. Однако, по мнению Гаспарова, в моем рассуждении присутствует ряд недостатков. Анализируя первую посылку, он отмечает: «Ни немыслимость, ни невозможность перевернутого спектра не могут служить решающим тестом для существования квалиа» (там же). С этим можно согласиться. Однако первая посылка моего рассуждения является условной, в ней не утверждается и не отрицается существование чего-либо, а лишь говорится, что чем бы ни были квалиа, если они обладают определенным свойством, а именно внутренней присущностью сознательным состояниям, то сценарий перевернутого спектра мыслим.

Далее Гаспаров пытается показать проблематичность первой посылки, утверждая, что возможно одновременное наличие того, что квалиа обладают свойством внутренней присущности, но ситуация



инверсии спектра немыслима. Соглашусь, что такая ситуация возможна, когда феноменальные качества необходимым образом связаны с телом. Однако в таком случае про эти качества уже нельзя сказать, что они характеризуются сущностным образом исключительно своим феноменальным измерением. Если они необходимым образом связаны с телом, то у нас нет оснований считать их чем-то принципиально иным, отличным от физических или функциональных свойств этого тела.

Объяснение феноменальных свойств, понятых таким образом, не является трудной проблемой. Последняя появляется тогда, когда мы рассматриваем феноменальные свойства как связанные случайным образом с телом. Сторонники нередуктивного физикализма, или, иначе, дуализма свойств, как раз стремились с помощью аргументов от представимости (*conceivability*), к которым следует отнести и аргумент от инверсии спектра, продемонстрировать, что феноменальные свойства связаны случайным образом с физическими и функциональными процессами нашего организма. Очевидно, что таким образом понятие феноменальных свойств, квалиа, если они внутренне присущи нашим сознательным состояниям, должны предполагать мыслимость перевернутого спектра.

Обращаясь к анализу второй посылки моего рассуждения, Гаспаров пытается продемонстрировать, что «витгенштейновские рассуждения не способны обосновать немыслимость перевернутого спектра» (с. 33). Как отмечает Гаспаров, «философ однозначно не отрицал возможность перевернутого спектра в любой форме». Действительно, он обсуждал сценарий интрасубъективной инверсии спектра. Следует признать, что этот вариант инверсии спектра вполне мыслим и, очевидно, возможен. Однако этот сценарий сам по себе не позволяет сформулировать аргумент в поддержку существования квалиа как внутренних свойств ментальных состояний. Ведь трансформация сознательного опыта субъекта сопровождается его бихевиоральным, функциональным и, возможно, физиологическим изменением. Все это не позволяет показать, что феноменальные свойства являются чем-то нередуцируемым к физическим или функциональным свойствам организма.

Примечательность сценария интрасубъективной инверсии спектра заключается в том, что он, по мнению некоторых философов, к которым примыкает Гаспаров, позволяет нам продемонстрировать мыслимость сценария интерсубъективной инверсии спектра. Однако это не верно. В своей книге в главе «Инверсия спектра» я подробно объясняю, почему это так (см. с. 152–183). Здесь же я лишь кратко отмечу следующее. Шумейкеру, на которого ссылается Гаспаров, не удалось показать, что в варианте интрасубъективной инверсии спектра речь идет именно об изменении квалиа. Изменяющиеся феноменальные

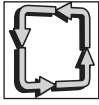


аспекты сознательного опыта субъекта могут пониматься как репрезентативные (интенциональные), функциональные состояния организма, не предполагающие наличия таких дополнительных свойств, как квалиа. Переход же от сценария интрасубъективной инверсии спектра к сценарию интерсубъективной инверсии спектра возможен, если мы уже изначально допустили существование квалиа. Собственно, такой маневр и совершает Шумейкер, тем самым делая ошибку *restitutio in integrum*. Другой философ, Нед Блок, пытаясь избежать этой ошибки, предлагает свой вариант этого перехода. Однако в результате он приходит к тому, что квалиа начинают представляться чем-то невыразимым, что, на мой взгляд, ставит под сомнение возможность сформулировать аргумент инвертированного спектра.

Как я полагаю, в конечном счете именно невыразимость (*ineffability*) квалиа препятствует мыслимости (*conceivability*) сценария интерсубъективного спектра. Поскольку эти понятия спровоцировали вопросы двух оппонентов – Игоря Гаспарова и Игоря Михайлова, проясню, каким образом я их себе мыслю. Понятие мыслимости я использую в том широком смысле, в каком оно функционировало в метафизических дискуссиях начиная с Декарта, обращавшего внимание на необходимость различать способность воображения и способность интеллектуального постижения чего-либо. Некая ситуация мыслима, если она непротиворечивым образом концептуально представима.

Почему сценарий инвертированного спектра немислим? Мы вполне можем вообразить эту ситуацию, как и любой другой фантастический сюжет, но пытаясь представить ее концептуально, мы сталкиваемся с противоречиями. Например, мы знаем, что понятие «красный» обозначает объекты – цвета, у которых мы можем выделить такие свойства, как яркость, насыщенность, светлота, тон и проч. Характеристики цвета вполне доступны объективному измерению и интерсубъективному обсуждению, они позволяют нам построить таблицы цветов, сверяясь с которыми, мы можем приходить к консенсусу в нашей совместной деятельности. Ни о каких таких свойствах квалиа мы не можем говорить. Чтобы не впадать в противоречия, допуская, что объекты, обозначаемые термином «красное», и обладают, и не обладают определенными характеристиками, мы должны признать, что лишь условно обозначаем квалиа цветовыми терминами. По сути это и означает, что квалиа невыразимы.

Игорь Михайлов задается вопросом, что означает эта невыразимость, но фактически абзацем ранее он сам же и отвечает на этот вопрос, говоря, что «слова выражают не квалиа, а социально сертифицированные способы различения свойств» (см. с. 37). Я бы не сказал, что цветовые термины выражают сертифицированные способы различения свойств, скорее они предназначены для выражения самих свойств, цветовых качеств объектов, но я согласен с Михайловым



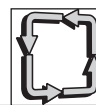
в том, что они не выражают квалиа. Иначе говоря, мы не можем сказать ничего конкретного об этих феноменальных свойствах, употребляя лишь условно для их обозначения цветовые термины. Все это и означает, что квалиа оказываются исключенными из контекстов, в которых мы могли бы работать с ними как с конкретными объектами, например обсуждать их природу и строить сценарий инвертированного спектра. Именно это я хотел продемонстрировать, используя идеи позднего Витгенштейна, при этом я не предполагал примыкать ни к какой теории значений: ни к верификационистской теории, ни к теории употребления (use theory).

Обосновывая то, что феноменальные свойства не являются внутренне присущими свойствами сознательных состояний, я не предлагаю отказаться от объяснения феноменальных характеристик сознания, открытие которых считаю важнейшим достижением Декарта. Я также не предлагаю объяснять их с позиций редукционизма. Почему-то именно таким образом проинтерпретировали мой подход Игорь Михайлов и Максим Беляев, хотя я отмечал, что вопрос о редукции не является тем основным вопросом, на который мы должны ответить, решая проблему сознания. Собственно, я не делаю акцента на том, каким именно образом мы будем вырабатывать натуралистическое объяснение феноменальных свойств.

Существуют различные натуралистические подходы к объяснению сознания, и не все они предлагают редуцировать сознание к физическим характеристикам. Например, мы могли бы интерпретировать сознательные состояния как поведенческие (бихевиоризм) или функциональные (функционализм) характеристики организма, отличные от его физико-химических свойств. Тот подход, который предлагаю использовать я, можно обозначить как репрезентационизм, или интенционализм. Оба названия используются как взаимозаменяемые в современной аналитической философии сознания. Репрезентационизм вполне совместим с натурализмом. Более того, он также совместим с функционализмом.

Игорь Михайлов задается вопросом: каким образом полученный на предыдущем этапе вывод, что феноменальные свойства являются внешними реляционными свойствами сознания, имплицитно концепцию интенциональности? (с. 39). На это можно ответить следующим образом. Прежде всего я не получал на предыдущем этапе вывод о том, что феноменальные свойства являются внешними реляционными свойствами сознания. Поскольку подобную интерпретацию с Игорем Михайловым разделили также Сергей Левин и Максим Беляев, проясню этот момент подробнее. Мой вывод состоял лишь в том, что квалиа не являются свойствами, внутренне присущими сознательным состояниям. Но из этого автоматически не следует, что они являются реляционными свойствами этих состояний. Такое следова-

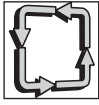




ние отсутствует в том случае, если квалиа вообще не существуют. В конце своей статьи Сергей Левин высказывает предположение о том, что «относительно квалиа стоит занять антиреалистическую позицию и не считать, что в реальности им соответствует что-либо, кроме принятых языковых конвенций» (с. 46). Отчасти я с ним согласен. Относительно квалиа я занимаю антиреалистскую или, правильнее сказать, элиминативистскую позицию. Квалиа, понятые как свойства сознательных состояний, не существуют, а то, что мы пытались обозначить термином «квалиа», является реляционными свойствами репрезентируемых объектов.

Левин предлагает рассматривать эти свойства как диадические, асимметричные реляционные свойства. Это очень интересный подход. В таких терминах я не думал о феноменальных свойствах. Действительно, хотелось бы рассматривать квалиа не только как свойства репрезентируемых объектов, но и как нечто, хотя бы отчасти характеризующее сознательные состояния организма, в которых этот объект схватывается. Однако, несмотря на то что реляционные свойства репрезентируемых объектов говорят нам что-то и о самих сознательных состояниях, их вряд ли можно рассматривать как характеристики этих состояний. Характеристика «быть высоким» вполне применима к репрезентируемому нами человеку, но не понятно, в каком смысле она могла бы быть также свойством самой ментальной репрезентации. (Думаю, что подобный пример можно использовать и для возражения против адвербализма, на возможность которого указывает Гаспаров.) Поэтому для того чтобы объяснить связь этих свойств с самими интенциональными состояниями, я написал: «Обсуждение онтологического статуса интенциональных объектов и их свойств должно вестись, на мой взгляд, только с точки зрения того, каким содержанием обладает интенциональное состояние» (с. 28). Иначе говоря, свойства репрезентируемого объекта фиксируются как часть содержания сознательного состояния, но не как свойства этого состояния.

По мнению Левина, феноменальные качества, понятые как реляционные свойства, не обладают особым онтологическим статусом, поскольку они лишены каузальной эффективности. Этот вопрос требует отдельного внимательного рассмотрения. Однако *prima facie* кажется убедительным положение о том, что наши действия все-таки зависят от изменения содержания наших интенциональных состояний. Если это так, то реляционные свойства репрезентируемых объектов все-таки обладают определенным онтологическим статусом, пусть и в качестве части содержания ментальных репрезентаций. При этом не думаю, что термины «содержание» и «репрезентация» следует рассматривать как сокращение для описания каких-то иных естественных свойств и состояний. Способность одних физических объек-



тов репрезентировать другие объекты сама по себе является естественным феноменом, который не нуждается в редукции к чему-то иному.

Поскольку возможность натуралистического объяснения сознания, на которую я указываю, ставится под сомнение в статье Максима Беляева, остановлюсь отдельно на этом моменте. В своей статье Беляев задается вопросом о том, «способствует ли понимание реляционного характера феноменальных свойств натуралистическому объяснению сознания» (с. 49). Он отвечает на этот вопрос отрицательно, объясняя это следующим образом: «Невозможность объяснения в данном случае связана с тем, что (1) сознание как вид сущего не мыслится одновременно объективно и субъективно существующим и (2) тела (вещи, материальные объекты, физические существа и т.п.) не обладают интенциональностью» (с. 51). На мой взгляд, оба пункта являются спорными.

Если субъективный тип бытия сущего, о котором пишет Беляев, не предполагает введения дуализма естественного и сверхъестественного, то следует признать, что сознательные состояния, наделенные такой характеристикой, как субъективность, являются естественными феноменами. Это значит, что мы можем не только представлять субъективные состояния как объективно существующие естественные феномены, но и пытаться объяснить их природу с объективистской позиции. В каком-то смысле, конечно, вся полнота знания о том или ином сознательном состоянии будет нам недоступна, в этом и проявляется специфика субъективности. Знание из перспективы первого лица, доступное субъекту сознательного состояния, всегда будет оставаться для нас недостижимым. Однако это не отменяет возможности объяснить природу сознательного состояния, если оно является естественным феноменом. Для объяснения важно, чтобы мы могли дедуцировать утверждения об этом феномене из других утверждений о естественных процессах. Иначе говоря, препятствием для натуралистического объяснения будет наличие разрывов в нашей онтологии, а не эпистемологии.

В своей статье я пытаюсь указать, каким именно образом нам следует мыслить субъективность, чтобы выработка натуралистического объяснения сознательных состояний была возможной. Для этого я использую распространенное в современной философии сознания понимание сущности субъективности посредством понятия перспективы первого лица. Я не ввожу это понятие для обсуждения интенциональности, как это пытается представить Игорь Михайлов, скорее наоборот, примыкая к Сёрлу, я пытаюсь прояснить понятие перспективы посредством введения понятия интенциональности. Понятие же интенциональности позволяет нам увидеть субъективность как естественный феномен. Оспаривая второе положение Беляева, я полагаю,

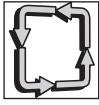


что любым физическим объектам можно приписать интенциональные состояния. Это хорошо показано, например, в работах Деннета. В этом нет ничего удивительного, если мы понимаем, что интенциональные состояния – это репрезентативные состояния и, очевидно, один физический объект способен представлять уникальным образом иной объект или какую-либо ситуацию. Некоторые философы, соглашаясь с этим, уточняют, что речь здесь идет о производной интенциональности, подлинными же интенциональными состояниями можно приписывать лишь существам, наделенным сознанием. Такой взгляд, например, отстаивает Сёрл. Не вникая сейчас в суть этого спора, отмечу лишь, что и философы последней группы, оставаясь на позициях физикализма, или – шире – натурализма, рассматривают интенциональность как состояние физических тел, прежде всего мозга.

Отвечая на замечания Михайлова, я хотел бы отметить, что в своей статье я не примыкаю специальным образом ни к какой особой теории интенциональности и не разбираю проблемы, возникающие в связи с этим понятием в философии сознания и философии языка. Я не делаю этого просто потому, что статья посвящена совершенно другой теме. Интенциональности я понимаю в самом общем смысле как способность представлять специфическим образом некий объект, и то, что я использую термин Сёрла «аспектуальность», чтобы подчеркнуть эту специфичность репрезентации, не значит, что я всецело разделяю его концепцию. Единственная особенность моего понимания интенциональности заключается в экстерналистской интерпретации интенционального объекта. При такой интерпретации никакого удвоения реальности не происходит.

Интенциональный объект не некая фикция «в моей голове», это наделенный различными свойствами реальный объект в момент, когда мы его репрезентируем. Те свойства этого объекта, о существовании которых имеет смысл говорить в контексте ее взаимодействия с другим объектом, мы обозначаем как реляционные свойства. Как я уже отмечал, можно думать об этих свойствах, как, впрочем, и о самом объекте, как о содержании репрезентативного состояния, помня, что понятием содержания мы фиксируем просто «факт особого отношения к миру» (с. 28). Не думаю, что нам следует проводить параллели между подобным элиминативистским и экстерналистским взглядом и локковским учением о вторичных качествах, что, по-видимому, подразумевает Михайлов.

Однако все же, по мнению Михайлова, мой «текст свидетельствует, что “на пути к объяснению сознания” мы вместе с Декартом и Локком едва лишь прошли его начало» (с. 40). Такая оценка выглядит странно. В п. 5 и 6 статьи я всего лишь воспроизвожу в самом общем виде репрезентационистский (интенционалистский) подход к объяснению сознания. Подобный подход лежит в самом основании когни-



тивной науки. Очевидно, когнитивная наука в XX в. значительно продвинулась вперед в понимании сознания и это стало возможным прежде всего благодаря критике идеи картезианского театра и расчистке феноменалистических джунглей британского эмпиризма. Однако в последней четверти XX в. перед когнитивной наукой вновь возникла проблема объяснения природы феноменальных данностей – квалиа (в первой половине XX в. этим термином обозначались свойства чувственных данных, сенсibiliа). Предлагаемое мной решение проблемы сознания, квалиа, новизна которого состояла в оригинальной постановке вопроса (внутренне присущи ли квалиа сознательным состояниям?), как я полагаю, должна способствовать нашему движению вперед в понимании психики по пути, уже проложенному когнитивной наукой. В последнем п. 7 статьи я даю свое понимание того, в каком направлении мы можем двигаться по этому пути с учетом тех трансформаций, которые происходили в самой когнитивной науке (коннекционистский поворот 1980-х, эннактивистский поворот начала 2000-х). Это ни в коем случае не выводы из статьи, скорее это общие размышления о действительно важной проблеме, стоящей перед когнитивной наукой и философией сознания.



# E XPERIENCING THE PRESENT<sup>1</sup>

**Yuri Balashov** –  
Department of  
Philosophy, University  
of Georgia, Athens, GA  
30602, USA. E-mail:  
yuri@uga.edu



I had excruciating back pain last night. The next day I went to a spa and the pain disappeared. Psychologically speaking, my pain is gone. Where is it, speaking ontologically? A-theorists have an easy time here (more or less). But B-theorists who think that persons persist by enduring are in trouble. Why am I finding myself at this particular time, with this particular set of experiences, rather than at numerous other times, with different experiences, despite the fact that all times are on the same ontological footing and I am wholly present at all of them? I argue that the Puzzle of the Experience of the Present is a real challenge for B-theorists, and the best way to deal with it is to adopt the stage view of persistence.

**Key words:** temporal experience, time, endurance, exduration, present.

## 1. Past Pains and Present Pleasures

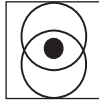
I had excruciating back pain early last night. The next day (which was, fortunately, a Saturday) I went to a spa, and the pain was gone. I started feeling warm pleasure spreading through my body as I stretched during the massage. I am still there, enjoying the experience.

Psychologically speaking, my pain is gone. Where is it, speaking ontologically? Perhaps it is in the past. If it's there why is it not painful? Perhaps because *I* do not *find* myself *there* (i.e. at night), but *here* (i.e. at a day session in the spa). But if I endure (roughly: if I persist by being wholly present at multiple time-places) then I should find myself both here and there (and in a lot of other time-places). Why am I *finding* myself during the day, in the spa, and feeling pleasure and not at night, in the bed, and in pain?<sup>2</sup>

The presentist (someone who thinks only the present is real) has an easy time (no pun) explaining these facts: the last-night pain is simply no more, and non-exi-

<sup>1</sup> This paper revisits an older project (Balashov, 2005), where the Puzzle of the Experience of the Present was put in a rather unfocused form, entangled with irrelevant matters. Recent flood of work on temporal experience has helped me focus my ideas. I have especially benefited from Skow (2011; 2015) and am very grateful for his comments on the previous versions of this paper at the PTS@APA session in San Francisco, California (March 2013) and thereafter. My thanks to the audience for pressing me on many points. Special thanks are due to Igor Gasparov, Mark Moyer and Jordan Lavender for their detailed comments on earlier drafts, and to Laurie Paul for her feedback on a recent draft. The result may or may not be up to their expectations; and I alone am fully responsible for any remaining defects.

<sup>2</sup> A skeptic may jump in at this point and say that the whole problem boils down to a simple confusion of the tensed and tenseless uses of 'find'. I *tenselessly* find myself at all times at which I exist; but I *find* myself (*present tense*) only on Saturday. I beg the skeptic to read on.



sting pains cannot be painful. Moreover, the night isn't there either, so I cannot find myself then. I find myself in the present because there is no past (or future). Of course, my pain *existed*,<sup>3</sup> and the night *was* present. But the fact that my pain existed does not put it in the past (or anywhere else for that matter), just as the fact that pigs possibly fly does not put flying pigs anywhere, for the modal actualist.

The moving spotlight theorist (roughly: someone who thinks that all times exist, but one of them is objectively privileged) needs to do more work to show that past pains are not painful. She needs to say that pains become painful only when the spotlight of the moving present shines on them. Maybe she can stop here. Or maybe she can say more about the shining mechanism.<sup>4</sup> But the idea is clear.

The real trouble awaits the eternalist (someone who thinks that all times exist and no time is objectively privileged) who also wants to be an endurantist. Or so I shall argue. In a nutshell, the argument is very simple. I am wholly present at night, and am also wholly present during the day. This immediately *connects* me with my pain as well as my pleasure. My relations to both are equally robust. No metaphysically significant feature of my relation to my current pleasure is missing from my relation to my last night's pain. In that respect, my pain is ontologically on a par with my pleasure. But they are not on a par as far as my experience goes. I am enjoying the pleasure, but I am not having the pain. Why? This raises a problem for the combination of eternalism with endurantism (EE). But this is a combination many philosophers want to have.<sup>5</sup>

## 2. The Puzzle and a Solution

Put this way the argument could be easily challenged; and rightly so. The perspective sketched in the previous paragraph presents an atemporal "view from nowhere." See Fig. 1. But our phenomenal perspective is inherently temporal; we can only experience things from a certain vantage point in time. See Figs. 2a and 2b, which represent such perspectives. To fix ideas, let us call the first, atemporal perspective *external*, and the second, inherently temporal one, *internal*. When the external and internal perspectives are properly distinguished it becomes clear that I can have my experiences only one at a time, not holus-bolus.

<sup>3</sup> Some presentists would analyze it as follows: it was the case that (my pain exists).

<sup>4</sup> For illuminating recent discussions (again, no pun intended) see [Zimmerman, 2005; Skow, 2015] and references therein.

<sup>5</sup> Here is a partial list of eternalist endurantists: [Johnston, 1987; van Inwagen, 1990, 2000; Mellor, 1998; Rea, 1998; Haslanger, 2003]. As stated here, the problem has much in common with Arthur Prior's famous "Thank Goodness that's over" puzzle, but is more general in scope pertaining to a whole variety of present temporal experiences, including but by no means restricted to differential attitudes towards the past and the future, such as relief, anticipation, or regret.

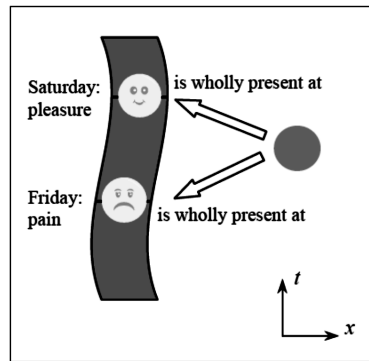
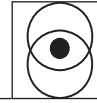
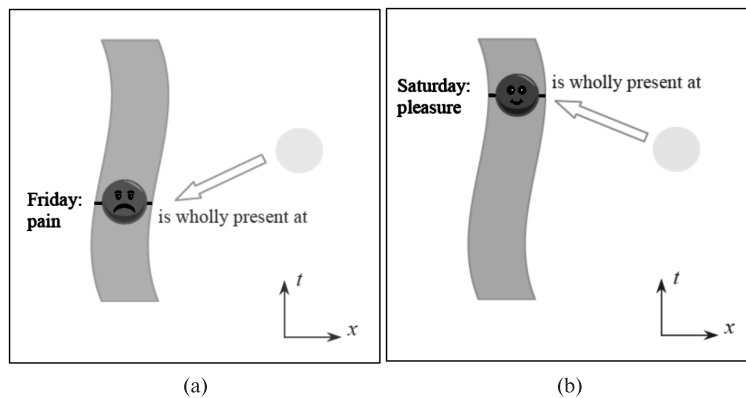


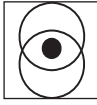
Fig. 1. The external perspective. The shaded area on the left represents my path in spacetime



Figs. 2a and 2b. The internal perspectives

This point should certainly be granted, but the argument can be refined [Balashov, 2005: 299ff]. Since temporal experiences can only be had one at a time the EEist should not fear being saddled with an impossible combination of pain and pleasure.<sup>6</sup> There is simply no internal perspective, in which this occurs. The real task is to explain why I am finding myself during the day, in the spa, and having pleasure, to the *exclusion* of pain, rather than at night, in the bed, and having pain, to the *exclusion* of pleasure. In short, why am I having the internal perspective of Fig. 2b rather than that of Fig. 2a? More carefully: why do I have an *illusory belief* that I am having the *exclusive* internal perspective of Fig. 2b, rather than an *illusory belief* that I am having the *exclusive* internal perspective of Fig. 2a? More gene-

<sup>6</sup> One should be careful here. For one thing, some experiences are known to combine pain and pleasure. More important, even for strictly incompatible types of pain and pleasure, there may be a difference between *having pain at t* and *having pleasure at t* and *having pain and pleasure at t*. I will not discuss these issues.



rally, why am I finding myself at this particular time, with this particular set of experiences, to the exclusion of others, *rather than* at any of the numerous other times, with different experiences, to the exclusion of others, despite the fact that all these times are on the same ontological footing and I am wholly present at all of them?

Here is another, slightly different take on it. The relation between external and internal perspectives is supposed to be *explanatory*. The external perspective of Fig. 1 is a true ontological picture of the block universe with endurance. This picture is required, among other things, to explain the appearances represented by Figs. 2a and 2b. There are two closely related but different explanatory questions to be addressed: (i) Why do I never experience an impossible combination of pain and pleasure, despite being connected to both (by the arrows depicted in Fig. 1)? (ii) Why do I experience pleasure rather than pain? Answering (i) is easy and a pre-requisite for (ii). I never experience pleasure and pain together because no arrow in Fig. 1 connects me to both; instead each connects me to a different experience, and also to a different time. These connections are invoked to explain (ground, stand behind) my having pain on Friday and pleasure on Saturday. They can only be invoked one at a time. But why *this* time rather than *that*? There are many times at which I am wholly present and have experiences; but there is *only one* “me” – on the EE view, that is. Invoking the arrows immediately helps with (i). But it also brings with it the need to break the ontological parity between my pleasure and my pain, and between Friday and Saturday, in order to explain my illusory belief in their disparity. And there is nothing in the eternalist block universe capable of doing this job.

The real question, then, is the “rather than” question, not the “why not all at once” question.<sup>7</sup> Why pleasure *rather than* pain? And why Saturday *rather than* Friday? Call questions of this sort *external questions*. The Puzzle of the Experience of the Present (PEP) for EE arises when external questions are taken seriously. But should they be taken seriously?

Framing the issue this way suggests two different courses of action. One is to take external questions seriously, acknowledge that PEP is a real puzzle, read ‘finding’ and ‘having’ in ‘finding myself at this particular time’ (*rather than* at some other time) and ‘having this particular set of experiences’ (*rather than* other experiences) in a way that makes the problem real; but deny, contrary to the appearances, that I am finding myself, in the relevant sense, *only* at this particular time, without, however, slipping back

<sup>7</sup> As far as I know, the difference between these two types of questions was first pointed out by Frederick Ferré [Ferré, 1972] and Ronald Hoy [Hoy, 1978] in the context of debating Grünbaum’s thesis of the mind-dependence of becoming. In a very balanced response to Ferré’s critique of Grünbaum’s position, Hoy sketched a view that is very close to my own [Balashov, 2005 and the present paper] and to the view favorably discussed by Skow in [Skow, 2011; 2015]. Thanks to Brad Skow for drawing my attention to these earlier sources.





into the “holus-bolus” pseudo-problem – without saying that I am finding myself at many times “at once.”<sup>8</sup>

The other course of action is to deny that PEP is a puzzle in the first place. Instead, it is a pseudo-problem based on a confusion. The confusion does not consist in mistakenly attributing to me a mix of pain and pleasure (see above). Rather it lies in using a misleading and tendentious language to create a false impression that the whole story about my temporal experiences involves more than is represented in Fig. 1, in which a single subject is connected to multiple times and multiple experiences. Pinpointing the confusion should make it clear that there is nothing more to the story.

I consider this skeptical response to PEP in §3 below. In the remainder of the present section, I want to sketch a rather radical *dissolution* of the puzzle about my pain and pleasure – assuming it to be a real puzzle, pending more considerations pro and con.

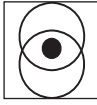
The best way to dissolve PEP<sup>9</sup> is to assign temporal experiences from different times to numerically different temporal *stages* of me. Each stage finds itself at a particular time and no others, and has the experiences it does at that time and no others, in exactly the same direct and non-relational manner as any other experience-bearing stage of me. In particular, my night stage finds itself exclusively at night, in bed, and in pain, and my day stage finds itself exclusively during the day session in the spa and enjoying pleasure. This is what allows me to say that I find myself, in the same simple way, at both times, with both experiences; but not “at once.”

Of course, this is not *so* simple. To save the phenomenal appearances which uniquely privilege my daytime experiences, we need to add that, in having the pleasant experience, I am *not aware* of having the painful one, and vice versa. This move to resolve the obvious tension involved in saying that I find myself at night, having pain, and *also* find myself in the middle of the day, having pleasure, and that in having one, I am not aware of having the other, is parallel to that made by Parfit and others to resolve the tension involved in fission scenarios<sup>10</sup>. In both cases the solution comes at

<sup>8</sup> Note that ‘finding’ has a distinctly subjective flavor here; we have already assumed that I am objectively and wholly present at multiple times.

<sup>9</sup> Explored in [Balashov, 2005; Skow, 2011: §6; 2015]. Skow’s discussion also draws useful parallels with Prior’s “Thank Goodness that’s over” problem (see note 5 above), while also noting some important differences (see, in particular, [Skow, 2011: 383–384, note 1]. For systematic presentations of stage theory (see [Sider, 2001: §5.8; Hawley, 2001]).

<sup>10</sup> In such scenarios, one of my physical and/or psychological descendants, Lefty, has an illusory belief that he is uniquely me and is not aware of the existence of Righty who has the same mistaken belief. The lack of the relevant awareness, which is crucial to explaining away the false beliefs, can be illustrated even in more down-to-earth split-consciousness scenarios. See, e.g., a vivid description in [Parfit, 2008: 361–362]. Although a parallel between the temporal case and Parfit-style cases is of great help in pursuing this approach, Skow does not mention it. The parallel is fully exploited in [Balashov, 2005]. See also [Bayne, 2010], which debates various forms of “static disunity” (i.e. disunity *at a time*) manifested in unusual experiences, including the “split-brain syndrome,” such as *representational disunity*, *access disunity*, and *phenomenal disunity*. It should be noted that Bayne’s analysis is firmly grounded in real-life clinical cases, does not extend to thought experiments or personal identity *across* time, and is intended to support the phenomenal unity thesis.



the price of downgrading the relation between experience-bearing items from strict numerical identity to “continuity with the right sort of cause.” This commitment to what we now call the *stage theory* of persistence is, at best, rather implicit in the earlier work on personal identity and fission, but it figures prominently in contemporary discussions.<sup>11</sup>

The Puzzle is thereby eliminated – essentially by undercutting one of its presuppositions, the idea that the Friday as well as Saturday experiences belong to a single subject. The external question loses its force, because each external perspective now generates a single internal perspective (see Fig. 3).

But there is a cost: we are left with the picture of the “fragmented self” familiar from the earlier literature on fission and related matters. More needs be said about it. The main stumbling block to overcome is to spell out the exact relationship between *me* and *my stages*. A popular but wrong way of doing it is to say that I exist only on Saturday (more precisely, at some moment on Saturday) and *someone else* exists on Friday, and then let the chips fall where they may.<sup>12</sup> I am aware of the ideological pressures pushing one in this direction. They come from drawing close analogies with the modal case, but the outcome strikes me as deeply dissatisfying, for “someone else” has an equal claim to be *me*. This claim is much stronger in the temporal case than in the paradigm modal cases. I think we should begin by saying that there is a sense in which I am *all* my stages, but not “holus bolus”; perhaps I am all of them “severally but not jointly”; and *then* let the chips fall where they may<sup>13</sup>.

<sup>11</sup> See, in particular [Sider, 2001: 201–205].

<sup>12</sup> Cf. [Skow, 2011: 380ff].

<sup>13</sup> “Not holus bolus” and “severally but not jointly” are of course metaphors that may or may not be useful in formulating a better version of the stage theory of the self. But at the very least, they are useful in pushing one away from another version of four-dimensionalism, *perdurantism*, on which there is a sense in which I *am* all my stages “holus bolus” or “jointly.” If that is the case then only a *part* of me viz. my Friday *temporal part* is in pain, and another part – my Saturday *temporal part* is enjoying pleasure. What then explains my belief that I (not a part of me) am enjoying the pleasure, rather than suffering from pain? This belief may be illusory, but the perdurantist is at a further disadvantage in dealing with it than the stage theorist. In one respect, the perdurantist’s predicament here is similar to the EEist’s predicament. The latter boils down to the need to explain why I (a single temporally *unextended* subject) am viewing the universe from the Saturday, rather than the Friday perspective, despite the fact that I (the self-same subject) am *wholly* present on both days. The perdurantist predicament boils down to the need to explain why I (a single temporally *extended* subject) am viewing the universe from the Saturday, rather than the Friday perspective, despite the fact that I (the same-self subject) am *partly* present on Friday, as well as Saturday. On both views, and in difference from the stage view, temporally-restricted experiences are had relationally or derivatively, not directly. One might say that the perdurantist still has an upper hand over the endurantist, because only the former can say that temporally-restricted experiences are had directly by *something*. Cf. Lewis’s response to Lowe [Lewis, 1988: 66]. But I agree with Skow [2011: 383] that I can be confident that pains and pleasures are had directly (and not relationally or derivatively) by something only in virtue of being confident that they are had directly by *me*. Although there is more to be said about the perdurantist option, I put it aside in the sequel and focus on the pros and cons of EE and the stage view.

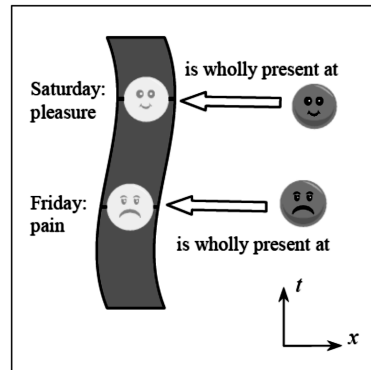
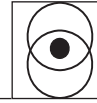


Fig. 3. Temporal experiences in stage theory

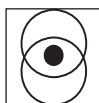
This cuts out more work. But it is good work; for it is needed to alleviate the above-noted tension inherent in saying that I find myself at night, having pain, and *also* find myself during the day, having pleasure. Ignoring this tension would give the emerging solution a flavor of triviality, while there is something strikingly non-trivial about the experience of time. At the end of the day, it seems trifling to be told that my Saturday stages have my Saturday experiences, and my Friday stages have my Friday experiences. But it is not trifling to be told that it is Saturday rather than Friday. I think this feeling is right. There is something non-trivial about the present experiences, and their non-triviality cannot be purchased at the price of something trifling. But there is something refreshingly non-trivial in the solution too: the insistence that I am *all* my stages – severally but not jointly, along with the claim that, in having my Saturday experiences, I am not aware of having my Friday experiences. I am not aware of them, and even of the very existence of my Friday stages, in the same sense in which Lefty (one of my fission descendants) may not be aware of the contemporaneous experiences, and even of the very existence, of Righty.

To sum up, the non-trivial part of the solution lies in rethinking the “I” part of the Puzzle, not the time part. And there seems to be no way around the Parfit road.<sup>14</sup>

### 3. External and Internal Questions

But the eternalist who is not enthusiastic about the stage view of persistence, or even about the need to take a stance on the issue of persistence

<sup>14</sup> Developing a more satisfying stage-theoretic account of persons consistent with this line of thought will have to wait for another occasion. For some very preliminary stabs at it (see [Balashov, 2005: 304–306]).



when dealing with temporal experience, may have a totally different reaction to the Puzzle of the Experience of the Present (PEP). She may simply refuse to see any puzzle here. Recall that PEP boils down to the following:

*The Puzzle of the Experience of the Present (PEP):*

Why do I find myself at this particular time, with this particular set of experiences, *rather than* at numerous other times, with different experiences, despite the fact that all these times are on the same ontological footing and I am wholly present at all of them?

Specifically:

Q1: Why am I having pleasure rather than pain?

Q2: Why is it Saturday rather than Friday?

Here are some alternative answers to these questions:

A1.1: Because the massage had its effect.

A1.2: Because the neurons in my brain were caused to fire in a pleasure-instantiating way.

A2.1: Because the sun has already crossed the date line.

A2.2: Because my calendar shows Saturday (and my calendar is never in error).

These answers look sensible. But they suggest that the questions are entirely non-puzzling for EE. What is going on here?

As stated above, Q1 and Q2<sup>15</sup> are ambiguous. They can be read as *external* questions, or as *internal* ones. External and internal questions are not different in their formulation; this is why they are so easily confused. They are different in terms of the background that stands behind them. The background for asking Q1 and Q2 in the external mode is the perspective of Fig. 1, while the background for asking the corresponding internal questions is the perspective of Fig. 2a or 2b. Perhaps one could say that an internal question is a question asked *from* a particular internal perspective (e.g. the Friday or the Saturday perspective), while a corresponding external question is about the *obtaining* of a certain internal perspective, as opposed to another such perspective.

PEP turns on taking external questions seriously. The Objection, on behalf of EE, is that such questions should not be taken seriously.

*Objection:*

External questions should not be taken seriously because (i) they make no sense at all, or (ii) if they do make sense, it is only because they are confused with the corresponding internal questions.

The Objection formally presents a dilemma. But it is best viewed as a two-pronged strategy. The main plan is (ii), while (i) is a back-up plan that can be activated if (ii) is in danger. Responding to (ii) is, I believe, a necessary prerequisite to dealing with (i). But I also think that an adequate response to (ii) goes a long way towards disarming (i).

<sup>15</sup>And a host of other questions: Why am I hearing the bell strike six times? Why am I seeing strange looking clouds pass by? Why is it so hot? Why is the summer almost over?



I have seen and heard the Objection in many forms, in and out of print. They all seek to locate the confusion between the external and internal questions in some more obvious and familiar confusion or oversight; for example, in forgetting that the B-theory of time is committed to the semantical reduction of tense; in a confusion between the first- and third-person perspectives; in the inability to appreciate the full significance of indexicals; in a failure to cope with self-locating beliefs and attitudes *de se*; in overlooking the presence of unarticulated constituents in Q1 and Q2, and/or a failure to see the analogy between the PEP and the problem of temporary intrinsics. I will return to the latter in the next section. But instead of rehearsing the responses already given to some of the former,<sup>16</sup> I want to illustrate a uniform strategy underlying such responses by applying it to a recent and particularly clear statement of the Objection due to Bradford Skow.

#### 4. Discharging Explanatory Tasks

To recap, the Puzzle of the Experience of the Present (PEP) for EE arises from reading

Q1: Why am I having pleasure rather than pain?

Q2: Why is it Saturday rather than Friday?

as external questions. The response at this point is to refuse to do so and read them as internal questions instead. Here is a good way to put the refusal [Skow, 2011: 376]<sup>17</sup>:

I am having pleasure rather than pain because my pleasure occurs on Saturday and it is on Saturday that I lie on the massage table wondering about my current experiences. There is no more to my pleasure than its occurrence at the time I do the wondering.

But this provides only a temporary relief. For the same could be said about my pain:

I am having pain rather than pleasure because my pain occurs on Friday night, and it is on Friday night that I lie in my bed wondering about my pain.

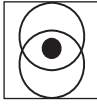
If this is the full story about my pain and pleasure then I am inclined to put my foot down and say that I want to know why I am having pleasure *and* wondering about it, rather than having pain *and* wondering about it. There are two episodes of wondering on Friday and Saturday, one of them painful the other pleasant. We still want to know what appears to break the parity between them in the EE perspective (Fig. 4):

Skow has suggested<sup>18</sup> that something so far unaccounted for does break the parity: *this* episode of wondering occurs only on Saturday. But I submit

<sup>16</sup> See [Balashov, 2005: §4].

<sup>17</sup> I took the liberty to adapt Skow's statement to my case.

<sup>18</sup> In personal communication.



that, rather than breaking the parity, a move to demonstratives or other devices of direct reference only enriches its content. *This* episode of wondering occurs only on Saturday. And *that* episode of wondering occurs only on Friday (Fig. 5).

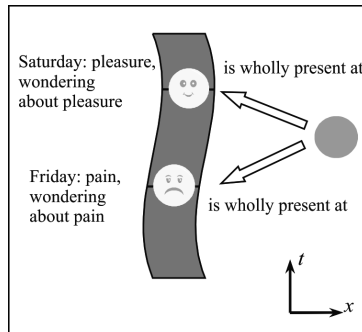


Fig. 4. Wondering about pain and pleasure in the EE universe

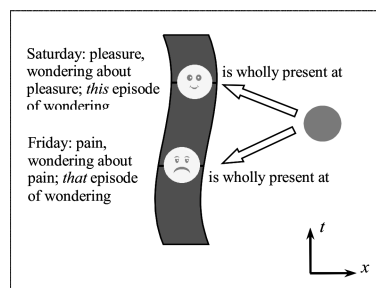


Fig. 5. Two episodes of wondering in the EE universe: *this* and *that*

What we want to know, however, is why *this*, and not *that* episode is *occurring*.

Recall that the Objection's plan A is to argue that putative external questions are in fact internal questions in disguise. That this is not so – that external questions are interestingly and importantly different from the corresponding internal questions – is already suggested by the drastic difference in the perspectives that stand behind them, those of Fig. 1 versus Figs. 2a and 2b. The above discussion shows that external questions tend to reemerge from scrutiny, sometimes becoming “thicker” as a result, and always distinct from the corresponding internal questions.

But perhaps this is just an appearance. Maybe what emerges and re-emerges this way makes no sense at all. The Objection's plan B is to condemn external questions as by-products of misleading theoretical schemas that fail to represent anything real. Let us take another look at Figs. 1, 4 and 5. What do they represent? Not anyone's phenomenal perspective; those are represented by Figs. 2a and 2b. Then what? Specifically, what does the



empty face on the right represent? And what are the arrows connecting it to the smiling and sorrowful faces on the left?

All I can say in response is that Figs. 1, 4 and 5 are my best way to illustrate the metaphysical picture of endurance in the block-universe setting inspired, among other sources, by a famous description due to David Lewis [Lewis, 1988: 66–67]:

Imagine trying to draw a picture of two different times,  $t_1$  when I sit and  $t_2$  when I stand. You draw two circles, overlapping because I exist at both times so you want to draw me in the intersection ... as a mere dot or shapeless blob, draw a line labelled ‘bent-at’ from me to the  $t_1$  circle and a line labelled ‘straight-at’ from me to the  $t_2$  circle.

It should be clear that temporary experiences could be assimilated to temporary properties. To say that I have pain on Friday is just to say that I have a certain mental property on Friday.<sup>19</sup> But instead of drawing times as overlapping and myself inside their common area (I cannot imagine how this could be done in my figures) I drew myself on the side and followed the custom of associating times with slices of four-dimensional volumes representing the paths of persisting objects in spacetime. And instead of labeling my arrows ‘pain-at’ and ‘pleasure-at’ I chose to assign them a more modest role of representing only the *multilocation* relation characteristic of EE, while placing the metaphysical burden of grounding pain and pleasure squarely on the relata. Nothing turns on this particular choice, and other semantical and metaphysical options are readily available.<sup>20</sup>

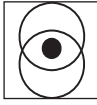
This leaves the empty face on the right. Apparently, it cannot represent the phenomenal self in a way the filled faces on the left do. What does it represent then? Well, perhaps it represents a *metaphysical self*.<sup>21</sup> After all, Fig. 1 is a metaphysical picture, and it is normal for metaphysical pictures to contain some metaphysical items.

Although I reject endurance, this picture makes *prima facie* sense to me. But this is a picture that stands behind external questions. They are sensible because they are explanatory. An external perspective is supposed to explain the origin of the internal perspectives from which I view the world. (In Figs. 2a and 2b this is illustrated by pushing the external perspective to the background and bringing the internal ones to the front.) Recall that there are two parts to this task: (i) explain why the internal perspective from which I view the world is a perspective of a single time, and not of many times at once; (ii) explain why the perspective from which I am viewing the

<sup>19</sup> What else could it be? This entails no broader commitments. A mental property may be a physical property; or it may not.

<sup>20</sup> Time may be involved in modifying a *property* or an *experience* itself, or the *having* of them; and the latter may be understood as the modification of the *instantiation* relation, or of the *obtaining* of a certain general *state of affairs*. See [Haslanger, 2003] for a nearly complete review of the available options.

<sup>21</sup> The term due to Kit Fine [Fine, 2005: 312ff] who uses it for a different, if not totally unrelated, purpose. The connection between what he does and what I do, however, is remote.



world is the internal perspective of a *particular* single time, *rather than* some other single time.

EE provides the resources to deal with (i): it is *because* I am wholly present on Saturday (in the external perspective) that I have my Saturday experiences (in an internal perspective); the same can be said about my Friday experiences. But EE has no resources to explain why I am viewing the world from *this* internal perspective *rather than* that one. All attempts to deflate questions of this sort fail because, given EE, my mode of existence in spacetime does not privilege any particular internal perspective. To wit, a single “I” gets associated with many “nows.” PEP results from a specific manner in which the personal perspective interacts with the temporal perspectives in the block universe with endurance: a single personal perspective appropriates many temporal perspectives indiscriminately.

I do not think the same sort of problem arises for objects that lack temporal experiences – pokers, candles, and so forth. For this reason, I do not think the “problem of temporary intrinsics” is a real problem for EE. Relations to times are special and do not deprive temporary intrinsics of their intrinsicity.<sup>22</sup> And there is nothing odd in drawing a poker outside of time, connecting it with different times and treating all such connections on a par. But when it comes to me and my experiences I want to resist this way of drawing them – not because experiences could not, in principle, be at bottom relational,<sup>23</sup> but because standing in a one-many relation to all my experiences from different times cannot explain why I *believe* that I am having *this one*, to the exclusion of the rest, *rather than that one*, to the exclusion of the rest; and relatedly: why I am having the *illusory belief* that I am present exclusively on Saturday, *rather than the illusory belief* that I am present exclusively on Friday.

The origin and tenacity of this belief must be adequately and convincingly explained, but its claim to veracity needs, of course, to be undermined. This can be accomplished by eliminating the recurrence of a single term (i.e. *me*) from the experience-generating relations, thereby preserving full parity among what goes on, metaphysically, at different times. There is an unfortunate mismatch between the personal and temporal perspectives in EE: one’s perspective on the world is naturally associated with a particular self, but not with any particular time. There is, accordingly, a pressing need to make the temporal perspective as compartmentalized as the personal perspective already is. The stage view is uniquely fit to address such explanatory concerns.

<sup>22</sup> Here I side with Hawley [Hawley, 2001: §1.4], Haslender [Haslanger, 2003], and others.

<sup>23</sup> To say that *S*’s temporal experience *E* is “at bottom relational” is to say *E*’s metaphysical status is adequately characterized by the description ‘*S* has *E* at *t*’ requiring a place holder for time, in order to be “saturated.” This could be embellished in various ways; but the crucial idea is that the description involves a three place predicate, perhaps a complicated one.





## References

- Balashov, 2005 – *Balashov Yu.* Times of Our Lives: Negotiating the Presence of Experience // *American Philosophical Quarterly*. 2005. Vol. 42. P. 295–309.
- Bayne, 2010 – *Bayne T.* *The Unity of Consciousness*. Oxford : Oxford University Press, 2010.
- Ferré, 1972 – *Ferré F.* Grünbaum on Temporal Becoming: A Critique // *International Philosophical Quarterly*. 1972. Vol. 12. P. 426–452.
- Fine, 2005 – *Fine K.* Tense and Reality // *K. Fine. Modality and Tense*. Oxford : Clarendon Press, 2005. P. 261–320.
- Haslanger, 2003 – *Haslanger S.* Persistence Through Time // *M.J. Loux and D.W. Zimmerman (eds.) The Oxford Handbook of Metaphysics*. Oxford : Oxford University Press, 2003. P. 315–354.
- Hawley, 2001 – *Hawley K.* *How Things Persist*. Oxford : Clarendon Press, 2001.
- Hoy, 1978 – *Hoy R.* Becoming and Persons // *Philosophical Studies*. 1978. Vol. 34. P. 269–280.
- Johnston, 1987 – *Johnston M.* Is There a Problem about Persistence? // *Proceedings of the Aristotelian Society (Supplement)*. 1987. Vol. 61. P. 107–135.
- Lewis, 1988 – *Lewis D.* Rearrangement of Particles: Reply to Lowe // *Analysis*. 1988. Vol. 48. P. 65–72.
- Mellor, 1998 – *Mellor D.H.* *Real Time II*. L. ; N.Y. : Routledge, 1998.
- Parfit, 2008 – *Parfit D.* Divided Minds and the Nature of Persons // *Peter van Inwagen and Dean Zimmerman (eds.) Metaphysics: The Big Questions*. 2nd ed. Oxford : Blackwell Publishing, 2008. P. 361–362.
- Rea, 1998 – *Rea M.* Temporal Parts Unmotivated // *Philosophical Review*. 1998. Vol. 107. P. 225–260.
- Sider, 2001 – *Sider T.* *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford : Clarendon Press, 2001.
- Skow, 2011 – *Skow B.* Experience and the Passage of Time // *Philosophical Perspectives*. 2011. Vol. 25. P. 359–387.
- Skow, 2015 – *Skow B.* *Objective Becoming*. Oxford : Oxford University Press, 2015.
- van Inwagen, 1990 – *van Inwagen P.* Four-Dimensional Objects // *Noûs*. 1990. Vol. 24. P. 245–255.
- van Inwagen, 2000 – *van Inwagen P.* Temporal Parts and Identity Across Time // *The Monist*. 2000. Vol. 83. P. 437–459.
- Zimmerman, 2005 – *Zimmerman D.* The A-theory of Time, the B-theory of Time, and ‘Taking Tense Seriously’ // *Dialectica*. 2005. Vol. 59. P. 401–457.



## 0 ДВУХ УРОВНЯХ ЭПИСТЕМОЛОГИИ СОЗНАНИЯ<sup>1</sup>

**Разеев Данил Николаевич** – доктор философских наук, профессор, заведующий кафедрой философии науки и техники Санкт-Петербургского государственного университета  
E-mail: drazeev@yahoo.de

В статье рассматривается вопрос о стратегиях познания субъективности. Автор полагает, что стратегия познания сознания подразделяется на два уровня. Базовый уровень эпистемологии субъективности, заданный философией, рациональной психологией и наукой Нового времени, различается на ф-эпистемологию (эпистемологию физических фактов, включая нашу телесность) и м-эпистемологию (эпистемологию ментальных событий). В XX в. в процессе натурализации сознания и открытия неинвазивных средств в изучении работы мозга происходит формирование следующего уровня эпистемологии субъективности. Автор предлагает различать на втором уровне р-эпистемологию (эпистемологию редуктивного типа, которая сводит или идентифицирует деятельность сознания с серией нейронных, сложных электрохимических и физических процессов в природе) и н-эпистемологию (эпистемологию нередуктивного типа, которая придает сознанию особый статус в природе).

**Ключевые слова:** эпистемология сознания, психофизическая проблема, натурализм, редуктивный и нередуктивный физикализм.

## 0 IN THE TWO LEVELS IN THE EPISTEMOLOGY OF CONSCIOUSNESS

**Danil Razeev** – DSc in Philosophy, professor and Head of the Department of Philosophy of Science and Technology, Saint Petersburg State University



The article considers different approaches to investigating consciousness. The author claims that consciousness study can be divided into two different epistemological levels. The first level of the epistemology of subjectivity was developed by philosophers and rational psychologists during the period of Early Modern Time, and can be divided into ph-epistemology, i.e. the epistemology of physical facts including our bodies, and m-epistemology, i.e. the epistemology of mental events. In the XX century, by naturalizing mind and inventing new non-invasive tools in investigating our brain, the epistemology of subjectivity went to the next level. The author considers contemporary movements in philosophy of mind working on the second epistemological level: eliminative materialism, property dualism, functionalism and so-called mysterianism. The author suggests dividing the second level into r-epistemology, i.e. the reductive epistemology that reduces and identifies our consciousness activity with a set of neuronal, electro-chemical and physical processes in nature, and n-epistemology, i.e. the non-reductive epistemology that considers consciousness as an independent phenomenon in nature.

**Key words:** epistemology of consciousness, mind-body problem, naturalism, reductive and non-reductive physicalism.

В данной статье предметом обсуждения будет не онтология сознания (т.е. не то, что такое сознание), а его эпистемология (т.е. то, каким образом мы познаем или изучаем наше же собственное сознание).

<sup>1</sup> Статья подготовлена в рамках реализации проекта РФФИ № 13-06-00902 «Нейронный коррелят сознания: теоретико-философский анализ современных естественно-научных концепций сознания».

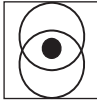


## I

Несмотря на то что уже среди самых первых образцов поэтических творений, философской и протонаучной мысли, древней медицины мы можем найти рассуждения о человеке как особом существе, о различии души и тела, о влиянии нашей телесности на изменение состояний сознания и т.д., существенное, так сказать, парадигмозадающее событие происходит, с моей точки зрения, в эпоху Нового времени, когда учеными и философами были разработаны принципиально различные эпистемологии (стратегии познания) физической и ментальной реальностей. Иными словами, для познания физического мира были установлены совершенно иные критерии, нежели для познания мира психического.

Для познания физических объектов было предложено применять так называемые количественные методы, благодаря которым физическая реальность поддавалась математическому исчислению, причем таким образом, чтобы можно было предсказать последующий шаг в ее изменении. Поскольку для такого предсказания одного лишь наблюдения за физической реальностью оказывалось недостаточно, то в науке Нового времени был изобретен специальный метод, а именно: метод искусственного создания условий, при которых физическая реальность раскрывала бы те закономерности, согласно которым она функционирует. Этот метод получил наименование эксперимента. Благодаря эксперименту ученые получили возможность выдвигать определенные гипотезы, которые впоследствии можно было подтвердить или опровергнуть опытным путем. Еще одной немаловажной характеристикой стратегии познания физического мира стала общедоступность перепроверки результатов предшествующих исследований. Таким образом, эпистемология физической реальности включала в себя несколько базовых принципов: наблюдение, гипотезу, эксперимент, а также доступность результатов познания в так называемой перспективе третьего лица.

Если окинуть общим взором эпистемологию (или, как уже было отмечено, стратегию познания) ментальных событий, то здесь господствующее положение получили совсем иные принципы. Базовым из них считалась интроспекция, или принцип самонаблюдения. Поскольку интроспекция предполагала наблюдение за ментальной жизнью конкретного субъекта, то сам этот субъект в эпистемологическом плане получал своего рода приоритетное или эксклюзивное право на наблюдение за своей психической жизнью. Для возможного обсуждения ментальной жизни нескольких субъектов предполагалась процедура фиксации переживаемых ими ментальных событий в языке, что требовало от субъекта, обладающего привилегированным доступом к событиям своей ментальной жизни, соблюдения критериев интеллектуальной честности при их фиксации и изложении.



Наконец, критерием истинности не могла выступать ни одна из ментальных цепочек какого-либо конкретного субъекта или даже их совокупность, а исключительно прообраз ментального, существующий совершенно независимым образом и находящий свое отражение в форме религиозного предания. Таким образом, эпистемология ментального предполагала: интроспекцию как базовый метод, привилегированный доступ субъекта к собственной ментальной жизни, так называемую перспективу первого лица, интеллектуальную честность, а также соотнесенность с религиозной истиной.

К чему же привело вышеуказанное различие в стратегиях познания физического и ментального? К своего рода расчленению живого субъекта на физический организм, который изучали медицина и экспериментальное естествознание, и на интеллектуальную сущность, или личность, с которой имели дело так называемые свободные искусства и науки (позднее гуманитарные науки). Медицина, следуя одной стратегии, была нацелена на изучение и лечение физического тела, функционирующего по физическим и биологическим законам, а свободные искусства и гуманитарные науки, следуя совсем другой стратегии, были призваны воспитывать душу, живущую по нравственным принципам.

Я предлагаю обозначить эпистемологию объектов физического мира (включая телесность человека) ф-эпистемологией, а эпистемологию ментальных событий (психики человека) – м-эпистемологией.

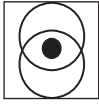
## II

К середине XX в. изучение субъективности перестает автоматически ассоциироваться только лишь с эпистемологией ментального (м-эпистемологией), для познания психической жизни субъекта начинают применяться принципы эпистемологии физического (ф-эпистемологии). Этот эпистемологический поворот связан с направлением в психологии, получившим наименование бихевиоризма, основы которого были заложены американским психологом Б.Ф. Скиннером (см., например: [Skinner, 1953]). Бихевиористы стали первыми психологами, программа которых состояла в том, чтобы объяснить психическую жизнь, руководствуясь принципами ф-эпистемологии, успешно применяемыми для научного познания различных феноменов природы, не прибегая к м-эпистемологии. С точки зрения бихевиористов, объективно, так сказать, из перспективы третьего лица, можно наблюдать лишь за поведением существа, обладающего психической жизнью. Соответственно любые гипотезы и эксперименты в области психологии и изучения субъективности, если они претендуют на научный статус, должны коррелировать с такого рода наблюдениями за человеческим поведением.



Представьте себе, что вы находитесь в цирке, где перед вами выступает известный фокусник, достающий из черного ящика один за одним какие-то предметы, которые просто не могли там оказаться, поскольку это противоречит законам физики. Какие у вас могут быть варианты узнать устройство черного ящика? Вы можете предположить, что черный ящик фокусника представляет собой аномалию, действующую по особым физическим законам. Но это вряд ли продвинет вас в объяснении устройства черного ящика. Вы можете положиться на рассказы фокусника, описывающего то, как устроен его черный ящик. Однако вряд ли стоит ожидать, что фокусник действительно раскроет вам секреты устройства своего магического ящика. Посещая одно представление фокусника за другим, вы также можете выстроить своего рода мифологию черного ящика, выявить определенные закономерности появления предметов из него и даже объяснить те случаи, когда отдельные ваши предсказания не сбываются. Все развитые вами теоретические конструкции будут разворачиваться в рамках м-эпистемологии. Для того чтобы выстроить их в парадигме ф-эпистемологии, необходимо сменить установку и переключиться на такое описание искомого феномена, которое сможет объяснить его исключительно в терминах физических событий. Объяснить феномен черного ящика в рамках ф-эпистемологии будет равнозначно способности повторить тот или иной фокус самому. Но как это сделать? Грубо говоря, предложенная бихевиоризмом стратегия состояла в том, чтобы отвести взгляд от черного ящика и внимательно следить за поведением самого фокусника, детально фиксируя все его физические движения, а затем научиться сопоставлять их с теми физическими эффектами, которые ассоциируются у нас с явлениями из черного ящика. Иными словами, бихевиоризм предложил перевести взгляд с самих ментальных событий на те физические проявления, которые коррелируют с ними в физической реальности. Программа бихевиоризма стала своего рода мейнстримом в области наук о психике в XX в.

Другим господствующим течением в изучении психической жизни в XX в. можно считать фрейдизм. Однако, как неоднократно справедливо замечал философ науки К. Поппер, теорию З. Фрейда нельзя считать научной в полноценном смысле этого слова [Popper, 1972]. Причиной этого служит своего рода всеядность фрейдизма. Под всеядностью имеется в виду черта теории Фрейда обратить любой факт, с первого взгляда противоречащий этой теории, в свою пользу. Теория всеядна, когда нет никакой возможности проверить ее истинность, поскольку нельзя сформулировать такую гипотезу, которая либо подтвердит, либо опровергнет центральные положения этой теории. Еще одной получившей широкую известность теорией сознания, созданной в рамках м-эпистемологии, можно считать феноменологию Э. Гуссерля. Ее создателю удалось добиться невероятных резуль-



татов, пользуясь исключительно методом самонаблюдения и рефлексии. Если бихевиоризм стал в XX в. господствующей парадигмой исследования психики в рамках ф-эпистемологии, то психоанализ и феноменологию можно считать известными методами изучения психики, сформулированными в рамках м-эпистемологии.

Вместе с тем, если окинуть общим взглядом развитие науки в XX в., необходимо отметить, что сознание в традиционном смысле этого слова, как совокупность ментальных событий, присущих определенному индивиду, не могло быть предметом изучения даже в рамках бихевиоризма как научной психологии. Как уже было отмечено, бихевиоризм занимался теми процессами, которые можно было наблюдать объективно, из перспективы третьего лица, а именно: поведением (выраженным в определенной последовательности физических действий) существа, обладающего психикой. Соответственно те процессы, которые происходят до пересечения своего рода границы их выражения в поведении субъекта, не подлежали и не могли подлежать никакому объективному изучению (хотя, нужно отметить, они могли быть реконструированы исходя из того, к каким поведенческим реакциям, наблюдаемым в физической реальности, они приводят).

Такое шаткое положение с изучением ментальных процессов в рамках ф-эпистемологии (а именно: опосредованное их изучение через изучение коррелирующего с ними поведения) сохранялось вплоть до появления в последней четверти XX в. нового научного направления, получившего наименование нейронауки. Основателем нейронауки считают испанского ученого и физиолога, лауреата Нобелевской премии по медицине Сантьяго Рамона-и-Кахала. Именно он еще в начале XX в. благодаря своим кропотливым наблюдениям выдвинул одну важную гипотезу, заложившую в дальнейшем новую парадигму в понимании работы нашей психической жизни [Ramón y Cajal, 1906]. С его точки зрения, нервные клетки не соединены друг с другом непосредственно, а разделены промежутками, пусть и невидимыми для невооруженного взгляда человека, которые впоследствии получили наименование синапсов. Важное следствие, которое вытекало из этой гипотезы, получившей затем подтверждение, заключалось в том, что нервная система живого существа не может считаться механизмом в традиционном понимании этого слова. Для того чтобы сигнал от одной нервной клетки перешел к другой, требуется ряд сложных электрохимических процессов. Иными словами, даже если тело живого существа есть механизм, то этот механизм устроен настолько многосложным образом, что он не может быть объяснен физическими законами базового (первого) уровня. Для объяснения работы нервных клеток требуется законодательство следующего (второго) уровня, а именно законы биологии. Идея Рамона-и-Кахала

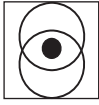


заклучалась в том, что, достигнув определенного уровня сложности, в организме формируется такая активность нервной системы (нейронная сеть), объяснение которой требует чего-то большего, нежели знания базовых физических законов, – требуется знание законов работы всей нервной системы организма, изучением которых должна заниматься отдельная дисциплина – нейронаука.

Технические возможности современной нейронауки позволили наблюдать и фиксировать процессы, происходящие в мозге живых существ, без непосредственного (инвазивного) вмешательства в его работу. Больше того, современные технические методы наблюдения и фиксации сложных физических процессов, происходящих в мозге, подняли вопрос о сознании и ментальном мире на совершенно иной уровень. Отныне проблема сознания обрела полное право войти в домен научного исследования в рамках ф-эпистемологии, поскольку ученые получили возможность не только теоретически реконструировать сложные процессы в мозге человека, но и наблюдать за ними в режиме реального времени. Раньше наука о психике вынуждена была руководствоваться только вербальными отчетами субъекта о происходящем в его ментальном мире, а также наблюдать за его (субъекта) поведенческими реакциями, теперь же в ее арсенале появился важнейший дополнительный уровень наблюдения – наблюдение за физическими процессами, происходящими в его мозге. Воспользуемся еще раз приведенной выше аналогией с фокусником в цирке: благодаря новейшим технологиям у ученых появилась возможность как бы в рентгеновских лучах наблюдать за тем, что происходит в самом черном ящике. Безусловно, «рентгеновский» взгляд сам по себе не в состоянии раскрыть всех секретов фокусника, однако он позволяет приблизиться к пониманию того, что делает фокусник, чтобы тот или иной фокус удался.

### III

Фундаментальный вопрос, который в наши дни стоит на повестке дня, заключается в том, возможно ли с помощью нейронауки в рамках ф-эпистемологии объяснить, что такое сознание, либо же для этого требуется отдельная научная дисциплина (дисциплина нового уровня), которая, не сводя сознание к функционированию нервных клеток (нейронов) мозга, будет способна объяснять его в парадигме ф-эпистемологии. Это сложный вопрос, который я связываю с эпистемологией сознания второго уровня (чуть позднее я вернусь к этому) и на который современная наука пока не может дать однозначного ответа. Вместе с тем существует несколько позиций по этому поводу в современной философии сознания. Они таковы: элиминативный материализм, дуализм свойств, функционализм и так называемый новый мистицизм.



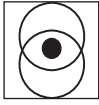
Элиминативный материализм (или элиминативизм) – это научная программа, согласно которой наука о психике должна быть выстроена исключительно на принципах нейронауки (см., например: [Churchland, 1986]). Речь идет о следующем. В ходе своего развития человечеством были выработаны определенные способы описания психических, или ментальных, процессов и событий. Совокупность такого рода описаний, исследовательских практик и выстроенных с их использованием теорий элиминативисты предложили обозначить термином «народная психология» (по аналогии с так называемой народной медициной – своего рода практическим знанием об использовании тех или иных природных средств для лечения заболеваний и недугов, которое вырабатывалось в течение многих тысячелетий людьми, проживающими в определенном регионе). Как современная медицина не имеет ничего общего с народной медициной, поскольку строится на научных основаниях, а не на обыденных представлениях и традиционных практиках обхождения с телом, так и научная психология, или наука о сознании, должна отбросить терминологический аппарат народной психологии и осуществить своего рода перевод всех значимых концептов психологии на язык нейронауки, что даст возможность наблюдать, проверять и опровергать гипотезы, т.е. работать в рамках ф-эпистемологии. Позиция элиминативизма, как можно заметить, достаточно радикальна, поскольку она не предполагает существования каких-то особых ментальных процессов. Существуют лишь физические, а вернее, сложные электрохимические процессы в мозге человека, нейронные процессы, которые и вписаны в единую физическую реальность. Все остальное – это лишь наши домыслы, которым нет места в научной психологии. Апеллируя к истории науки, элиминативисты указывают на достаточное число примеров, когда тот или иной процесс, пока он не был понят в рамках ф-эпистемологии, поначалу получал особый онтологический статус, но впоследствии благодаря тщательному научному изучению отвергался. Достаточно напомнить, что наука Нового времени заявляла о существовании эфира или флогистона как особых материальных силах в природе, однако спустя время стало понятно, что они были лишь выдумкой ученых, которые не могли, используя имевшийся в их распоряжении концептуальный и технический аппарат, найти верное физическое объяснение ряда физических феноменов.

Немного иную позицию занимают философы-физикалисты, которые придерживаются теории дуализма свойств (см., например: [Block, 2007]). Они считают, что физическая реальность, оказавшись или достигнув определенного состояния, может порождать такие физические свойства (так называемые эмерджентные свойства), которые нельзя полностью редуцировать к предыдущим ее состояниям как замкнутой физической системы. Именно таковы ментальные про-





цессы в природе. В качестве примера эмерджентного свойства можно привести описание того, как люди учатся плавать. Сначала вам очень долго кто-то объясняет, как нужно передвигаться в воде, держать наклон тела, двигать руками и ногами, делать вдох и выдох, задерживать дыхание, и т.д. В принципе подобного рода инструкции могут быть достаточно ясно прописаны и вы в состоянии вызубрить их и даже сымитировать, находясь на суше. Однако стоит вам оказаться в воде, как вы, несмотря на все знание и тренировку, можете утонуть. Потому что плавать – это что-то большее, нежели совокупность механических движений разными частями тела в соответствии с определенными правилами. В этом смысле плавание есть эмерджентное свойство, которое у кого-то из проинструктированных и натренированных на суше индивидов, помещенных в воду, может возникнуть, а у кого-то нет. Можно научиться удерживаться какое-то продолжительное время на поверхности воды, соблюдая определенную технику движений тела, но для того чтобы уметь плавать в строгом смысле этого слова, требуется что-то большее, а именно: своеобразное чувство, некая уверенность вроде той, когда мы говорим себе: «Да, я могу это!» Сторонники дуализма свойств полагают, что для объяснения ментальных феноменов нет нужды апеллировать к независимо существующей и действующей по особым законам ментальной субстанции, радикально отличающейся от объектов физического мира; однако в то же самое время они полагают, что не все свойства физических объектов могут быть редуцированы к базовым физическим законам. В сложных биологических системах в определенный момент их развития возникает нечто вроде фактора  $\phi$ , который просто невозможно разложить на базовые физические составляющие. Более того, невозможно заранее предсказать появление этого фактора. Из двух одинаково устроенных физических систем одна может обладать эмерджентным свойством, или  $\phi$ -фактором, а другая – нет. Наше сознание есть не что иное, как такой  $\phi$ -фактор. Это довольно деликатный момент, от понимания которого зависит понимание того, о чем вообще говорит одна часть философов, изучающих сознание, и за что их критикует другая часть философов, изучающих сознание. Как могут существовать две одинаковые физические системы, в одной из которых возникает определенное качество  $\phi$ , а в другой – нет? Если допустить, что существует только один мир – мир физических объектов и процессов, а сторонники дуализма свойств придерживаются такой точки зрения, и если этот физический мир функционирует согласно причинно-следственным связям, то крайне сложно найти адекватный ответ на поставленный выше вопрос. Сторонники дуализма свойств полагают возможным обойти эту проблему посредством постулирования своего рода онтологии, факта данности, но не двух независимых друг от друга субстанций, а одной физической субстанции, обладающей



дополнительным фундаментальным свойством –  $\phi$ -фактором. Сравните летящий в небе самолет и парящую в небе птицу. Возможно, с аэродинамической точки зрения, процесс их полета схож. Однако в отличие от самолета птица в известном смысле «чувствует» свой полет, чувствует, что такое летать. Даже если в недалеком будущем найдется умелец, который сможет изготовить такой совершенный в механическом отношении летательный аппарат, который будет в точности похож на птицу, и внешний наблюдатель не сможет отличить его от птицы, тем не менее фундаментальное отличие между ними будет состоять в том, что настоящая птица будет обладать уже упомянутым чутьем полета в отличие от ее механического дубликата. Это чутье птицы и есть дополнительное эмерджентное свойство, искомый  $\phi$ -фактор, который, правда, лишь онтологически постулируется и поэтому требует своего научно-философского объяснения.

Еще одна философская позиция по указанной проблеме – так называемый функционализм (см., например: [Fodor, 1981]). Функционализм тесно связан с идеей множественной реализации одной и той же функции на разных материальных носителях. Допустим, вы хотите узнать, сколько будет три плюс два. Для этого вы можете использовать два камня серого цвета и три разноцветных камня, можете использовать два камня и три кошки, можете использовать деревянные счеты, а можете карманный калькулятор или современный iPhone. Чем бы вы ни воспользовались, в результате операции получится число пять. Но при чем здесь сознание? Дело в том, что, с точки зрения функционалистов, работа сознания – это совокупность реализации определенных функций. Вопрос о конкретном материальном носителе для реализации этих функций как бы выносится за скобки. Имеется в виду следующее. Событие  $x$  может быть определено как в терминах материальных причин, так и причин функциональных. Это значит, что событие  $x$  сбывается в тот момент, когда для этого выполнен ряд условий. Трюк состоит в том, что эти условия могут быть выполнены на разных материальных носителях или в разных материальных системах. Одна и та же операция может быть реализована как в системе Windows, так и в системе Linux, несмотря на то что это разные языки программирования и что одна операция выполняется на стационарном компьютере фирмы Hewlett Packard, а другая – на портативном компьютере от корпорации Apple. Материальная составляющая не определяет правил реализации самой функции, вернее, существует множество материальных систем, на которых та или иная функция может быть исполнена. Таким образом, функционалисты полагают, что ментальные события могут реализовываться на разных материальных носителях (в том числе, но не исключительно, на нейронных!). Соответственно они отказываются редуцировать работу сознания к функционированию нейронов в мозге, поскольку последний



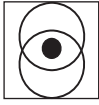
представляет собой лишь один из образцов материальной системы, на которой возможно разыграть работу сознания. Функционалисты отказываются описывать работу интеллекта в терминах причинно-следственных связей, построенных на анализе работы нейробиологических механизмов мозга. Они предпочитают описывать когнитивные функции вне их связки с нейробиологией.

Последнее из наиболее известных направлений в новейшей философии сознания – это так называемый новый мистерианизм, который представляет собой современную форму агностицизма (учения о недостаточности имеющихся в нашем арсенале способностей для адекватного познания реальности). Мистерианцы, будучи физикалистами, т.е. признавая сознание частью природы, полагают, что вопрос о том, что оно такое, в принципе не может быть решен в силу ограниченности тех познавательных средств, которые присущи нашей форме интеллекта (см., например: [McGinn, 1999]). Даже если нейронауке в будущем и удастся проследить все без исключения связи нейронов друг с другом в мозге того или иного человеческого индивида (напомню, что в мозге около 100 миллиардов нейронов и каждый из них может вступать в своего рода коммуникацию с 10 000 соседних нейронов), она не сможет полноценно ответить на вопрос о том, почему и как у него возникло сознание. Дело в том, что нейронаука в состоянии установить определенные корреляции между работой сознания и работой мозга, но не в состоянии выявить причинно-следственную связь, поскольку для этого требуется быть мета-существом, обладающим мета-сознанием. Открытие и детальное описание нейронного коррелята сознания, с точки зрения мистерианцев, не даст ответа на вопрос о том, почему и когда сознание возникло, а также каково его место в природе.

#### IV

Хотя все вышеуказанные позиции в современной философии сознания постулируют свою приверженность физикалистской установке (ф-эпистемологии), важным представляется вопрос о том, каков тот тип физикализма, которого каждая из указанных позиций придерживается. Именно в этом я вижу эпистемологическую проблему сознания второго уровня. Рассмотрим этот уровень более подробно.

Как именно сознание вписано в единую физическую реальность? Мой ответ на этот вопрос, если сформулировать его предельно кратко, заключается в тезисе о необратимости сознания. Этот тезис означает, что в случае сознания мы имеем дело с фактом такой структурной организации физической реальности, который невозможно свести к базовым физическим силам, на сегодняшний день известным науке, – сильной, слабой, электромагнитной и гравитационному взаимодействию. Я сторонник так называемого нередуктивного физика-



лизма, или, как я предлагаю это обозначать, н-эпистемологии сознания. Попробую прояснить, что это означает. Существуют различные способы объяснения разнородных феноменов, которые нельзя путать между собой. Представьте себе мир, в котором составляющие его феномены как бы разбросаны по различным этажам одного и того же здания. Несмотря на единство здания в целом и невозможность существования каждого последующего этажа здания без наличия предыдущего, каждый из них спроектирован и функционирует по своим собственным правилам. Это означает, что для объяснения конкретного феномена, находящегося, скажем, на третьем этаже, нам потребуется обратиться к правилам, по которым функционирует именно третий этаж здания, а не второй или первый. Допустим далее, что в таком гипотетическом мире на первом этаже играют в футбол, на втором в бридж, на третьем в шахматы, а на четвертом в компьютерную стратегию. Хотя онтологически в этом гипотетическом мире игра в шахматы будет основываться на игре в бридж и в футбол (поскольку, как уже было сказано, они составляют предыдущие этажи здания), эпистемологически у нас не будет возможности объяснить ход конем G1–F3, апеллируя к правилам игры в футбол на первом этаже, к правилам игры в бридж на втором или правилам компьютерной симуляции на четвертом этаже здания. Очень огрубляя и упрощая, скажу, что реальность представляется мне подобного рода четырехэтажным зданием. Правила игры первого этажа определяются Лейбницевым протовопросом о том, почему вообще существует нечто, а не ничто. На втором этаже здания господствуют законы, по которым функционируют органические создания. На третьем этаже установлены принципы существования сознающих организмов. На четвертом же этаже здания сознающие органические существа пытаются скалькулировать правила для искусственного интеллекта. Более того, еще не известно, не будет ли в здании когда-то надстроен какой-то новый этаж или даже несколько. Итак, здание в целом представляет собой сложную физическую систему и в данный конкретный момент мы не в состоянии полностью редуцировать (обратить) кибернетику к науке о сознании, которую в свою очередь невозможно свести к биологии, а последнюю к физике. Вот в каком смысле я говорю о необратимости сознания как феномене природы. Иными словами, я понимаю под сознанием такой биологический феномен, который эпистемологически невозможно полностью свести к чисто биологическим механизмам несмотря на то, что онтологически этот механизм является частью физического мира (мира природы).

Пользуясь вышеприведенной аналогией с четырехэтажным зданием, я продемонстрирую и точку зрения редуктивного физикализма, которую я называю р-эпистемологией (и не разделяю), согласно которой все существующее можно объяснить (редуцировать) к базовым



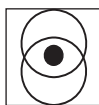
физическим принципам. Редуктивные физикалисты полагают, что на всех четырех этажах здания заданы одни и те же правила игры. Скажем, на первом этаже играют в дворовой футбол, второй этаж представляет собой футбольное поле в форме стадиона, третий этаж отведен под компьютерную симуляцию футбола, а четвертый – под сетевую игру в футбол двух или более компьютеров, в том числе без участия человека. Сторонники р-эпистемологии полагают, что структурно этажи могут отличаться друг от друга, но формально у них будет одна и та же онтология, заданная едиными правилами игры. Мне представляется, что такая позиция существенно упрощает физическую реальность как целое. Ее сторонники пытаются задать реальности не только единый синтаксис (единство базовых элементов, из которых состоит реальность на любых ее уровнях), но и единую семантику (унифицированную организацию этих элементов на всех уровнях).

Как я уже отметил, базовый тезис р-эпистемологии сознания, заключающийся в обратимости сознания, представляется мне спорным. Для доказательства тезиса обратимости сознания необходимо разработать такие научные теории, которые были бы в состоянии полностью объяснить (редуцировать) деятельность сознания в базовых физических терминах. Успешных теорий такого толка мне не известно. (Хотя попытки их создания имеют место в современной науке о сознании, в частности этому посвящены исследования С. Хамероффа и Р. Пенроуза, см.: [Hameroff, Penrose, 2014].)

Таким образом, в современной философии и науке, с одной стороны, господствует натуралистический подход к сознанию в рамках эпистемологии субъективности первого уровня (установка ф-эпистемологии), согласно которой сознание есть часть природного мира, а с другой стороны, сосуществуют две конкурирующих между собой эпистемические установки второго уровня, одна из которых (р-эпистемология) сводит работу сознания к базовым физическим процессам в природе, в то время как другая (н-эпистемология) рассматривает сознание как особый природный феномен, бытие которого невозможно полностью свести к принципам работы других ее элементов (н-эпистемология, разумеется, не означает наличия в существе, обладающем сознанием, какого-то сверхприродного, или трансцендентного, начала).

## V

Таким образом, я провел различие по типам и уровням эпистемологии субъективности. На первом уровне в стратегии познания субъективности, сформированной наукой, рациональной психологией и философией Нового времени, различимы м-эпистемология (эпистемология ряда ментальных событий) и ф-эпистемология (эпистемоло-



гия физических фактов, включая нашу телесность). На втором уровне познания, уже в ходе разворачивания программы натурализации сознания в XX в., возникли р-эпистемология (эпистемология редуктивного типа, которая сводит или идентифицирует деятельность сознания с серией нейронных, сложных электрохимических и физических процессов в природе) и н-эпистемология (эпистемология нередуктивного типа, которая придает сознанию особый статус в природе).

Разумеется, указанными различиями эпистемологические вопросы касательно стратегии познания сознания не заканчиваются, а только еще обозначаются вчерне. Последовательной программой разработки эпистемологии субъективности было бы решение следующих вопросов: 1) Имеет ли теоретико-познавательное значение различие второго уровня (различие на р-эпистемологию и н-эпистемологию) не только для физикализма, но и для ментализма, т.е. на базовом уровне м-эпистемологии? 2) Насколько будет отличаться второй уровень эпистемологии сознания в зависимости от того, какая из программ (ф-эпистемология или м-эпистемология) реализуется на первом уровне? 3) Могут ли выявленные отличия в стратегии познания на втором уровне оказаться эвристически значимыми для различных стратегий познания первого уровня, а также для общей стратегии познания субъективности? 4) Исчерпывается ли общая стратегия познания субъективности исключительно двумя вышеуказанными уровнями? Это вопросы открыты для дальнейшего изучения.

## Библиографический список

- Block, 2007 – *Block N.* Consciousness, Function, and Representation: Collected Papers. Vol. 1. Cambridge : MIT Press, 2007.
- Churchland, 1986 – *Churchland P.S.* Neurophilosophy: Toward a Unified Science of the Mind-Brain. Cambridge: MIT Press, 1986.
- Fodor, 1981– *Fodor J.* The Mind-Body Problem. Scientific American 244, 1981.
- Hameroff, Penrose, 2014 – *Hameroff S., Penrose R.* Consciousness in the Universe: A Review of the ‘Orch OR’ Theory, Physics of Life Reviews : Elsevier. 2014. N.Y. Vol. 11 (1). P. 397–398.
- McGinn, 1999 – *McGinn C.* The Mysterious Flame: Conscious Minds in a Material World. N.Y. : Basic Books, 1999.
- Popper, 1972 – *Popper K.R.* Conjectures and Refutations. The Growth of Scientific Knowledge. L. : Routledge, 1972.
- Ramón y Cajal, 1906 – *Ramón y Cajal S.* The Structure and Connexions of Neurons. Nobel Lectures. 1906. December. 12.
- Skinner, 1953 – *Skinner B.F.* Science and Human Behavior. N.Y. : Macmillan, 1953.



## К «ПРАГМАТИЧЕСКОЙ» И ИНТЕРАКТИВНОЙ ПАРАДИГМЕ ИССЛЕДОВАНИЯ ИНТЕРСУБЪЕКТИВНОСТИ

**Федор Алексеевич Станжевский** – ассистент кафедры философии Санкт-Петербургского государственного технологического института (Технического университета). Защищает диссертацию (PhD) по философии в Ягеллонском университете, Краков. E-mail: stanzh@mail.ru

Репрезентационизм и интернализм не в состоянии передать всю сложность деятельности сознания и мозга в ситуации интерсубъективности. Возникает необходимость другого подхода, который учитывает как деятельностный характер индивидуального сознания, так и автономию процесса взаимодействия по отношению к его участникам. В статье приводятся описание и интерпретация ряда экспериментов группы исследователей, которые изучают работу мозга при интерсубъективном взаимодействии исходя из деятельностной и интерактивной, а не «зрительской» модели взаимодействия.

**Ключевые слова:** интерсубъективность, взаимодействие, репрезентационизм, интернализм, совместное внимание, атрибуция ментальных состояний, наблюдение, участие.

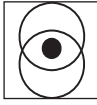
## TOWARDS A PRAGMATIST AND INTERACTIVE PARADIGM OF STUDYING INTERSUBJECTIVITY

**Fedor Stanzhevskiy** – a lecturer at the department of philosophy of the Saint-Petersburg State Technological Institute.

Modern cognitive sciences are dominated by the individualist paradigm of studying intersubjectivity. According to this paradigm, the individual mind precedes intersubjective interrelationships. The individual is a fully constituted participant of interaction. Interaction itself is nothing more than a sum of internal models in the minds and brains of the participants and as such it has no autonomy in itself. Consciousness reflects “inside” the outward objective reality. The representative theory of mind is one of the foundations of internalism and individualism about mind. In case of perception, a representation is construed as an internal mental state which represents the outward world. In case of action, a representation is an internal mental state that causally precedes outward behavior. These ideas imply a split between the individual mind and the world; a similar split exists between the interacting minds.

The article postulates the necessity to develop a truly interactive model of intersubjectivity. Such a model implies that interaction has reality of its own apart from internal models “in the heads” of its participants. This model would imply that individual minds are results of interaction no less than they are prerequisites for it. Based on data from developmental psychology, the author argues that individual minds are constituted in intersubjective interaction. The first stage of this interaction is dyadic interrelation whereby an infant and her mother constitute one dynamical system. The second stage is the so called joint attention, which engenders the triangulation of self-consciousness, other-consciousness and objectivity. Then some experimental neurobiological data are presented. These data have been obtained in an attempt to provide an ecological approximation to real interactions. These data were obtained and interpreted from a non-individualistic standpoint, which shows that the internalist model of cons-





sciousness is not the only one to serve as an experimental and theoretical paradigm in studying intersubjectivity.

**Key words:** intersubjectivity, interaction, representational theory of mind, internalism, joint attention, affordance, attribution of mental states, observation, participation.

В современных когнитивных науках господствует индивидуалистская парадигма исследования intersubjectivity, согласно которой индивидуальное сознание предваряет intersubjectivity, является уже «готовым» и вполне конституированным его участником. Предпосылка методологического индивидуализма многим исследователям кажется очевидной и не требующей доказательств; между тем этот предрассудок является лишь интеллектуальной привычкой. Одной из мощных (с точки зрения влияния, а не обоснованности) опор этой привычки является репрезентационизм. Согласно теории репрезентационизма (в некотором упрощении), сознание отражает «внутри» уже готовую «внешнюю» объективную действительность. В случае *восприятия* репрезентация – это внутреннее ментальное состояние, отражающее внешний мир. В случае *действия* репрезентация – это внутреннее ментальное состояние, каузально предваряющее действие «вовне». Настоящая статья не направлена на критику репрезентационизма как такового; мы хотим лишь указать на насущную необходимость смены парадигмы в исследовании intersubjectivity (переход от репрезентационистской модели к модели подлинно интерактивной). Поэтому ограничимся лишь кратким упоминанием тех проблем, к которым неизбежно приводит репрезентационистская модель.

Репрезентационизм резко отделяет «внутреннее» от «внешнего» и, следовательно, влечет за собой дуализм мир-сознание, поведение-сознание. Вместо того чтобы усматривать диалектическое сцепление между организмом и средой, на которое указывали еще американские прагматисты и русские нейрофизиологи, репрезентационизм проводит четкие границы между сознанием и внешним миром. Эти границы столь непреодолимы, что репрезентационизм совершает колебательное движение от реализма (напомним, что репрезентация отражает данную и заданную действительность) к солипсизму (если между репрезентацией и миром имеется разрыв, то почему бы репрезентация вовсе не утратить зависимость от мира – именно эту гипотезу подвергает критике Хилери Патнэм в работе “Brains in a vat”). Репрезентация как внутреннее ментальное состояние есть продукт гипостазирования процесса, «статизация» подлинной динамики, имеющей место между мозгом, телом и средой. Репрезентация – это статический снимок движения; стало быть, основывать изучение взаимодействия субъекта со средой и с другими субъектами на понятии репрезентации – значит судить о «Лебедином озере» или «Русском Гамлете», опираясь на ряд фотографий.





Однонаправленная причинно-следственная связь (при восприятии – снаружи вовнутрь, а при действии – наоборот) существенно упрощает действительную картину работы мозга, в котором процессы “top down” и “bottom up” дополняют друг друга и оказывают друг на друга взаимное воздействие. Так, Ален Берто [Berthoz, Petit, 2006] показывает, что при визуальном восприятии процесс «синтеза» визуальной картины уже в самых первых своих стадиях, в областях первичной зрительной коры VI, испытывает обратное воздействие со стороны предлобной коры и лимбической системы. Таким образом, нет смысла говорить о дуализме центра и периферии, а скорее о диалектике между ними. Именно этим объясняется воздействие на восприятие наших привычек, социальных и индивидуальных, а также нашей мотивации. Мозг не является пассивным органом, рецептором информации, напротив, мозг – это орган действия, непрестанно совершающий проекцию действительности. Потенциал готовности, наблюдаемый в области дополнительной двигательной коры перед движением, является тому примером: мозг проецирует планируемое движение до его совершения. Мозг не столько воссоздает действительность, сколько предваряет ее. Так, для хищника имеет значение не сама по себе адекватность воспроизведения движений своей жертвы: для эффективного прыжка важно рассчитать будущее местоположение добычи, предварить момент, который еще не настал.

Познание живого организма изначально деятельно, а восприятие всегда сопряжено с движением. Именно последняя идея лежит в основе теории кинестезиса, которую (возможно, независимо друг от друга) предложили Уильям Джеймс и Эдмунд Гуссерль. Кинестезис представляет собой неразрывное единство между восприятием и движением, которое делает возможным это восприятие. Для того чтобы «охватить взглядом» здание Генерального штаба, надо совершить ряд движений, как саккадических, так и мышечно-скелетных – повернуть голову и изменить направление взгляда. Полагают даже [O'Regan, Noe, 2001], что закодированное в головном мозге различие модальностей восприятия (например, слух и зрение) основывается исключительно на различии двигательных программ.

Методологический индивидуализм, основанный на репрезентационистском интернализме, как мы уже сказали, воздействует на исследования в области интересности. Приведем пример такого исследования, в котором посылка индивидуализма руководит интерпретацией эксперимента. В одном из своих исследований Julian Keenan et al. (2003) при помощи МРТ изучали возможную локализацию самосознания в головном мозге человека на примере распознавания испытуемым собственного лица. Во время сканирования субъектам показывали: (1) изображение их собственного лица, (2) лицо какой-либо известной личности, которую следовало индексировать,



и (3) компьютерную версию теста на распознавание выражения глаз (36 изображений человеческих глаз, выражающих различные ментальные состояния – серьезное, задумчивое и т.д.), причем субъектов просили распознать ментальное состояние человека по картинке.

В отличие от условия (2) при изображении собственного лица (1) наблюдалась достоверная активация в правой верхней, средней и нижней лобной извилинах. При выполнении теста на распознавание выражения глаз (3) активировались лобные области с обеих сторон, а также медиальная область лобной доли. При этом достоверная активация наблюдалась в средней лобной извилине с обеих сторон, левой верхней височной извилине и медиальной зоне верхней лобной извилины. Таким образом, области активации в случаях (1) и (2) перекрывались. Другими словами, области коры, активируемые при обработке информации о себе, в значительной мере совпадали с областями, активируемыми при обработке аналогичной информации о другом субъекте (условие эксперимента (1) – распознавание своего лица – было нацелено на исследование самосознания, а условие (3) – распознавание выражения глаз – на изучение понимания другого человека). Из наблюдения совпадения областей активации в обоих случаях авторы заключают, что способность моделировать и понимать сознание другого человека основана на способности обработки самореферентной информации, т.е. способность понимать самого себя более фундаментальна по сравнению со способностью понимать других людей. Однако как бы ни обстояло дело с истинностью этого тезиса, проблема заключается в том, что полученные экспериментальные данные никак его не обосновывают: все, что мы можем узнать из данного эксперимента, – это наличие корреляции (а не каузальной зависимости) между, так сказать, «обработкой самореферентной информации» и «атрибуцией ментальных состояний» другому субъекту, иными словами, между самопониманием и пониманием других. Тезис о каузальной зависимости, выдвинутый авторами, является плодом более или менее неосознанного допущения, а не результатом эксперимента, следовательно, этот тезис экспериментально недоопределен. Проблема в том, что исследователи проецируют заранее принятое допущение индивидуализма на результаты эксперимента.

Представляется, что существуют достаточные основания полагать, что индивидуальное сознание не предваряет взаимодействие, а, напротив, является его следствием. Причем это утверждение безоговорочно справедливо в отношении онтогенеза индивидуального сознания и с определенными очевидными оговорками – в отношении конкретного ситуативного взаимодействия между «зрелыми» субъектами.

Психология развития предоставляет достаточно материала для иллюстрации этого тезиса в отношении онтогенеза индивидуального

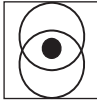


сознания. Схематически можно выделить несколько стадий развития интерсубъективности у детей. Немаловажно, что языковое взаимодействие должно с неизбежностью основываться на доязыковых формах интерсубъективности.

Первая стадия интерсубъективности включает в себя феномен имитации новорожденных: новорожденные младенцы в состоянии точно имитировать движения лица взрослых, приводя в соответствие зрительно воспринимаемые ими, но не ощущаемые движения лица с ощущаемыми проприоцептивно, но не видимыми жестами своего собственного лица. Это сцепление (весьма вероятно, врожденное, но, по-видимому, требующее актуализации в конкретном опыте) между проприоцепцией и зрительным восприятием представляет «межтелесную связь» (*intercorporality*, по выражению Мерло-Понти) между Я и другим. Возможно, именно это явление иллюстрирует врожденную и непосредственную, т.е. не выводимую в умозаключении, способность человека воспринимать другого человека именно как наделенного субъективностью, как равного ему субъекта.

Феномен диадической первичной интерсубъективности выходит за рамки простой имитации. Эта первичная интерсубъективность представляет собой ряд сенсомоторных (а лучше сказать – кинестетических) навыков, способностей понимать смысл движений, жестов, выражений лица другого человека (воспитателя), значение направления его глаз, а также намеренные действия в контексте взаимодействия лицом-к-лицу. Ребенок уже не просто имитирует, а реагирует на эмоциональные выражения лица воспитателя. Речь идет о подлинном взаимодействии, в котором ребенок и воспитатель взаимно регулируют эмоции друг друга. Этот динамический процесс взаимной регуляции превосходит сумму репрезентации эмоций воспитателя у ребенка и репрезентации эмоций ребенка у воспитателя – речь идет о подлинном, динамическом обмене, о *взаимодействии*. Этот эмоциональный обмен между воспитателем (матерью) и ребенком в определенном смысле расширяет границы индивидуального сознания обоих: имеет место диадическое расширение сознания [Tronick, 1998].

С точки зрения теории систем мать и ребенок – две самоорганизованные системы – обретают большую степень связности и сложности организации (в частности, большую сложность «состояний» организации мозга) именно в диадическом сцеплении. Индивидуальная система черпает большую степень сложности в диадически сцепленной системе; материнская поддержка регуляции эмоций у ребенка позволяет тому достичь более сложной организации мозга. Диадическая система обладает степенью сложности и связности, превосходящей эндогенные состояния матери и ребенка, взятые по отдельности, но индивидуальная система реконфигурирует свою организацию, интегрируя в нее «продукты» взаимодействия. Если говорить на языке



Выготского (с осторожностью, чтобы не путать разные парадигмы), происходит интериоризация того, что возникло в результате взаимодействия. Этот процесс напоминает процесс идентификации, описываемый психоаналитиками, который также знаменует собой интересубъективную подпитку индивидуального роста: Я первоначально возникает путем идентификации со значимыми другими. Стоит заметить, что диадическое расширение сознания во взаимодействии с другими, по всей видимости, далеко не ограничивается периодом развития, но (хотя и менее выражено) сопровождает любое интересубъективное взаимодействие.

Феномен совместного внимания характеризует так называемую вторичную субъективность (Trevvarthen, 1979) и является второй стадией развития сознания, непосредственно следующей за первичной интересубъективностью. Совместное внимание представляет собой ситуацию «триангуляции» (по выражению Д. Дэвидсона) Я–другой–объект в мире. Ребенок выходит за пределы восприятия намерений и эмоций другого человека, относящихся к его (ребенка) Я, теперь он схватывает то, как другие взаимодействуют с миром, используют вещи в мире, интенционально направлены на мир. Ребенок начинает понимать ситуации, контексты действия [Gallagher, 2005], разделять общий мир с другими. В этом смысле термин «внимание» несколько обманчив, поскольку не учитывает деятельного, телесного аспекта триангуляции. В ситуации совместного внимания ребенок не является зрителем, внутренне репрезентирующем визуальную картину: он деятельно и аффективно вовлечен во взаимодействие, которое само по себе является процессом, а не данностью.

Совместное внимание по своей природе связано с кинестезической природой восприятия вещей и с кинестезическими корнями интересубъективного взаимодействия. Как и первичная интересубъективность, совместное внимание не исчерпывается периодом развития; в частности, оно не только генетически, но и феноменологически проясняет феномен объективности. Здесь оказывается, что «объективность» объекта задается не только суммой моих кинестезисов по отношению к нему, но и перспективой (иной, чем моя) на этот объект другого субъекта. Перспектива же другого непосредственно понимается исходя из его моторных (кинестезических) возможностей действия и восприятия в мире. Исток моего понятия «объективной» реальности, таким образом, нужно искать не только в интермодальном (например, зрительном и тактильном) восприятии, но и в интересубъективной ситуации совместного внимания. В то же время и сама интересубъективность модифицируется благодаря опосредованности в объекте: теперь связь между Я и другим пролегает через разделяемую ими ситуацию, через предмет совместного внимания, который достигим не только для меня, но и для



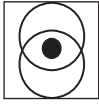
другого, лежит в пределе возможностей как моего действия, так и действия другого. Возможно, здесь получает развитие и приобретает полноту восприятие другого как субъекта, впервые актуализованное при имитации и вышедшее на новую ступень организации в первичной intersубъективности. Отныне другой субъект не только связан со мной эмоционально, но, как и я, является деятелем в мире, обладает возможностью действия, полем возможностей в данной ситуации. Разумеется, в треугольной ситуации совместного внимания модифицируется и самосознание, самовосприятие Я, которое получает возможность учесть перспективу другого на объект и на себя. Здесь Я участвует в организации другого, а другой в свою очередь участвует в организации Я.

Совместное внимание не только проясняет онтогенез самосознания, intersубъективности и объективности, но и служит моделью исследования «зрелого» intersубъективного взаимодействия как такового. Подобно тому как это было описано в случае диадической intersубъективности, треугольное взаимодействие обладает собственной динамикой, спецификой и автономией, не являясь простой суммой вкладов каждого из участников взаимодействия по отдельности. De Jaegher and Di Paolo (2007) описывают взаимодействие как координацию движений (мы бы сказали – кинестезисов) субъектов, при которой индивидуальное смыслополагание каждого из них испытывает на себе воздействие со стороны всего процесса взаимодействия. Последний обладает операционно замкнутой структурой, несводимой к суммированному поведению индивидов. Такое взаимодействие авторы называют участвующим смыслополаганием. Смысл активно порождается во взаимодействии, а индивид конституируется как деятель внутри процесса взаимодействия.

Из сказанного выше следует, что модель исследования взаимодействия должна основываться не на простом наблюдении субъектом другого и даже не на наблюдении им взаимодействия, а на деятельном участии во взаимодействии. По вполне очевидным причинам нейрофизиологические исследования такого взаимодействия связаны с определенными трудностями, откуда, однако, не следует неизбежность «сложить оружие»: во второй части настоящей статьи мы описываем попытку (пока одну из немногих) применить интерактивную модель исследования к изучению нейрофизиологии intersубъективности.

\*\*\*

Долгожданная смена парадигмы в исследовании социального сознания намечается в статье «К нейронауке второго лица» [Schilbach et al., in press], где авторы постулируют необходимость перейти от исследований субъекта в качестве наблюдателя к исследованию собственно *взаимодействия* между субъектами. В самом деле, ощущение



и понимание «ментальных состояний» другого человека в обычном контексте происходит не в виде наблюдения за другим, а во взаимодействии с этим другим человеком. Предварительные исследования при помощи методов нейровизуализации подтверждают существенное отличие нейронной активации в ситуации взаимности, характерной для социального взаимодействия, по сравнению с простым наблюдением.

Как мы уже видели, старая парадигма, все еще очень распространенная, негласно предполагает, что социальное сознание (т.е. способность познавать другого человека, его ментальные состояния, действия и т.д.) представляет собой такой способ обработки информации, который по существу является пассивным «поглощением» информации, предоставленной объективным и уже готовым к восприятию миром. Вопреки этому в новой перспективе восприятие является активным процессом со стороны организма, расположенного в среде и укорененного в ней, а отнюдь не абстрактно изолированного от нее. Эту характеристику восприятия передает труднопереводимый на русский язык термин “affordance” (его не совсем точно переводят как «возможность»), предложенный известным психологом Дж.Дж. Гибсоном. Среда предлагает организму определенные возможности (дерево как укрытие, вода как питье или как среда обитания, место охоты и т.п.). Восприятие животным этих возможностей зависит как от условий среды, так и от его собственной конституции – здесь имеет место сцепление организма со своей средой. Следовало бы говорить о применении такого понятия возможности к человеческой социальной среде – другой человек всегда предлагает возможность взаимодействия с ним как с другом, начальником, партнером и т.д.

Социальное сознание, интерсубъективность вообще, как правило, предполагает эмоциональную ангажированность и взаимодействие – в том смысле, в котором взаимодействующие субъекты взаимно отвечают на состояния или действия друг друга, вызывая цепь взаимных реакций друг на друга. В таком взаимодействии главную роль играет не теоретическое познание другого субъекта, но практическое ноу-хау, основанное на навыке взаимодействия и, напротив, формирующее этот навык. В контексте такого взаимодействия возникают общие, разделяемые субъектами намерения и мотивы, могущие приводить к совместным действиям и дальнейшему развитию взаимодействия. С точки зрения нейрофизиологии можно выдвинуть гипотезу о различии нейронной активации в зависимости от роли во взаимодействии, например является ли субъект инициатором или же отвечает на инициативу другого.

Прежде чем приступить к описанию экспериментов с ПЭТ и МРТ, иллюстрирующих это явление, отметим еще один важный



теоретический момент. Мы уже упоминали о диадической интерсубъективности, эмоционально связывающей двух субъектов. Эмоциональная вовлеченность в контакт с другим человеком имеет исключительную важность с самого детства и является источником смысла и познания в отношении как самого себя, так и другого человека. Это значение эмоциональной вовлеченности в контакте с другим человеком никогда не ослабевает: эмоции, по-видимому, играют огромную роль в восприятии и понимании других людей; эмоциональный ответ здесь гораздо важнее, чем интеллектуальные конструкции. Однако в повседневности общение между людьми имеет место в определенном контексте, в некоторой ситуации – ситуации совместного внимания. Повторим: речь идет о триадической интерсубъективности, в которой двое субъектов взаимодействуют и разделяют аффекты и установки в отношении третьего объекта или лица. Именно явление совместного внимания стало предметом исследования в экспериментах Л. Шильбах и др.<sup>1</sup>, призванных подтвердить правомерность новой парадигмы изучения социального сознания.

Шильбах и др. [Schilbach et al., 2006] исходят из предпосылки о различии нейрофизиологического базиса при исследовании: а) стороннего наблюдателя и б) субъекта, находящегося в реальной ситуации взаимодействия с соответствующей эмоциональной нагрузкой. Авторы пытаются моделировать более интерактивные и более экологически достоверные эксперименты, чтобы выявить нейронные корреляты социального сознания с точки зрения наблюдателя и с точки зрения субъекта взаимодействия. Кроме того, они задаются вопросом о том, могут ли два взаимодействующих мозга вместе предоставить больше информации о взаимодействии, чем их сумма по отдельности.

Шильбах с коллегами провел исследование при помощи функционального МРТ, в котором попытались выявить различие в нейронной активации в случаях, когда социальные стимулы направлены на субъекта (испытуемого) или на кого-то другого. Испытуемым предложили представить себе самих себя стоящими рядом с двумя другими людьми, которых они не видят. Перед ними находился третий человек, которого они могли видеть, – виртуальный персонаж, изображаемый на видео (рис. 1). Переменные эксперимента заключались в выражении лица виртуального персонажа (это были либо социально значимые, либо произвольные выражения лица), а также в направлении его взгляда (он смотрел либо прямо на испытуемого, либо в сторону, на одного из двух невидимых персонажей). Испытуемых просили оценить, ощущают ли они, что виртуальный персонаж намерен вступить в социальное взаимодействие.

<sup>1</sup> См.: Гранов [и др.] (в печати).

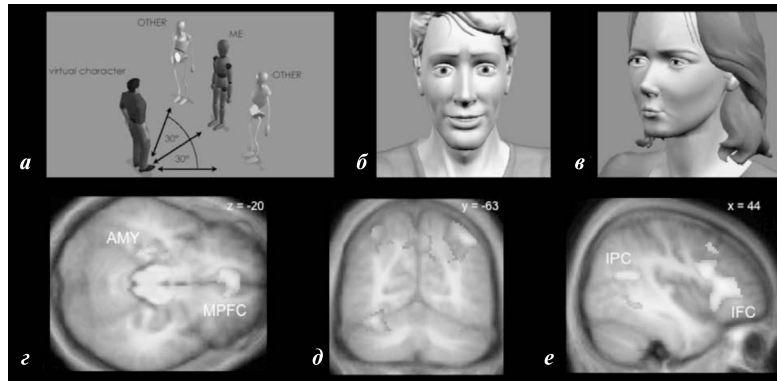
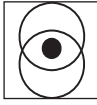


Рис. 1. Исследование с помощью МРТ [Schilbach et al., 2006]:  
*a* – виртуальный сценарий согласно инструкциям; *б* – социально значимое выражение лица, взгляд направлен на испытуемого субъекта; *в* – произвольное выражение лица, взгляд направлен на другого человека; *г* – нейронные корреляты восприятия испытуемым взгляда, направленного на него; *д* – нейронные корреляты восприятия испытуемым взгляда, направленного на другого человека; *е* – нейронные корреляты восприятия испытуемым произвольных движений лицевых мышц

Испытуемые без труда различали социально значимую и произвольную мимику; при этом большая концентрация внимания и большее эмоциональное вовлечение ощущались в случае, когда стимул (взгляд) был направлен на испытуемого. Результаты нейровизуализации (рис. 2) выявили достоверные различия в нейронной активации при восприятии социально значимой мимики в зависимости от того, был ли направлен взгляд на испытуемого или в сторону от него. В первом случае наблюдался дифференциальный рост нейронной активности в вентральной части медиальной области предлобной коры и на поверхности миндалевидного тела; в случае, когда взгляд виртуального персонажа был направлен на другого человека, активировались медиальная и боковая зоны теменной коры головного мозга.

В случае, когда взгляд виртуального персонажа направлен в сторону от испытуемого, последний выступает в роли наблюдателя; когда же взгляд виртуального персонажа направлен на самого испытуемого, имеет место начало социального взаимодействия. Сами авторы на основании описанных данных МРТ замечают, что в первом случае активируются области мозга, отвечающие за визуально-пространственную информацию, во втором случае стимулы, направленные на самого испытуемого, активируют области, связанные с обработкой эмоциональной и оценочной информации (миндалевидное тело и вентральную зону предлобной коры). Эти результаты подтверждают важную роль эмоций не только при восприятии эмоциональных состояний другого человека, но и при взаимодействии с ним; они обеспечивают телесный ответ на выражение, намерения и действия друго-



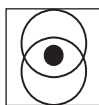


го и таким образом предоставляют возможности и мотивацию для действия и дальнейшего взаимодействия.

Кроме того, результаты МРТ показали, что при восприятии произвольной мимики активируются правая нижняя область лобной коры и нижняя теменная кора – области, обычно включаемые в систему зеркальных нейронов, активируемую при восприятии моторных интенций другого человека. Как бы ни обстояло дело с собственно зеркальными нейронами, сама активация указанных областей при восприятии действий другого человека убедительно проиллюстрирована при помощи методов нейровизуализации. Однако роль этой активации может быть интерпретирована по-разному. Обычно активацию областей коры головного мозга (ГМ), отвечающих как за выполнение, так и за восприятие действия, интерпретируют (в случае восприятия) как симуляцию действия другого. Однако возможна и другая интерпретация: активация этих областей может быть связана не просто с симуляцией действия другого, а с восприятием возможностей дальнейшего взаимодействия с ним исходя из этого только что совершенного действия. Иначе говоря, эта активация подготавливает мозг и организм в целом к ответу на действия другого, к комплементарному, дополняющему действию в процессе взаимодействия. При этом активируются моторные программы, позволяющие координировать собственные действия с действиями другого человека.

В другом исследовании Шильбах с коллегами изучал нейронные корреляты лицевой мимики, выявив сеть, состоящую из лицевой области первичной моторной коры, задней зоны поясной коры, гиппокампа и дорсальной части среднего мозга. Авторы делают вывод о существовании тесной связи между восприятием выражений лица и формированием аналогичных выражений лица у наблюдателя. Они заключают также, что ситуации, требующие эмоциональной вовлеченности и ведущие к некоему моторному ответу, активируют лобно-теменные зоны и срединные структуры коры ГМ, а также лимбическую систему ГМ.

Следовательно, первоосновная роль социального сознания – не просто интеллектуально познавать сознание других людей и их ментальные состояния. Напротив, познание другого человека основано на том ответе, который вызывают в нас его действия: «Когда другие обращаются к нам лично, восприятие их выразительного поведения, их мимики с нейрофизиологической точки зрения опирается на тесное сцепление восприятия и действия с обработкой аффективной и телесной информации, подпитывающей и поддерживающей подготовку моторного ответа как способа увидеть возможности для взаимодействия и ответить на них».



В своем исследовании М. Уилмс и др. [Wilms et al., 2010] изучали нейрофизиологические корреляты взаимности социального взаимодействия. Ведь взаимодействующие непосредственно воздействуют друг на друга, и это взаимное воздействие необходимо показать на уровне его нейронного базиса. Ясно, что современные аппараты МРТ на данный момент не позволяют представить всю полноту взаимодействия двух субъектов. Поэтому авторы смоделировали новый тип постановки эксперимента, основанный на интерактивном отслеживании движений глаз.

В эксперименте Уилмса испытуемый, помещенный в сканер, мог видеть видео, изображавшее виртуального персонажа; движения глаз испытуемого мониторились при помощи инфракрасной камеры. Движения глаз виртуального персонажа приводились в соответствие с взглядом испытуемого, а движения глаз последнего были направлены на «прошупывание» поведения другого, подобно тому, как это происходит в реальной жизни. Такая постановка эксперимента позволяет испытуемым ощущать, как их собственные движения глаз воздействуют на взгляд виртуального другого человека. Кроме того, эксперимент позволяет исследовать нейронные корреляты роли испытуемого в социальном взаимодействии – речь идет о роли инициатора или же реагирующего на инициативу. Это немаловажно при исследовании совместного внимания, когда мы направляем свое внимание на что-либо совместно с другим человеком, осознавая эту нашу с ним совместную направленность внимания. Ситуация совместного внимания возникает либо когда мы следуем за взглядом другого, либо когда сами направляем его взгляд.

Шильбах и его коллеги использовали представленную выше постановку эксперимента при помощи МРТ для исследования совместного внимания [Schilbach et al., 2010]. Испытуемые взаимодействовали с виртуальным персонажем во время нейровизуализации. Испытуемый либо инициировал ситуацию совместного внимания с персонажем, либо, напротив, следовал за взглядом этого персонажа, в обоих случаях фиксируя взгляд на предметах, показываемых на экране. В результате МРТ выявила активацию медиальной области предлобной коры и задней зоны поясной коры при совместной направленности внимания на объект независимо от роли – инициатора или реагирующего. Направленность взгляда на другой предмет, а не на тот, на который смотрел виртуальный персонаж, активировала боковую лобно-теменную сеть (рис. 2, в), которая связана с контролем внимания и движений глаз. В свою очередь медиальная предлобная кора и задняя поясная кора принадлежат к ментальной системе, описанной выше, и перекрываются областями, отвечающими за так называемый режим мозга по умолчанию (default mode), которые активируются в состоянии покоя ГМ при отсутствии внешних задач...

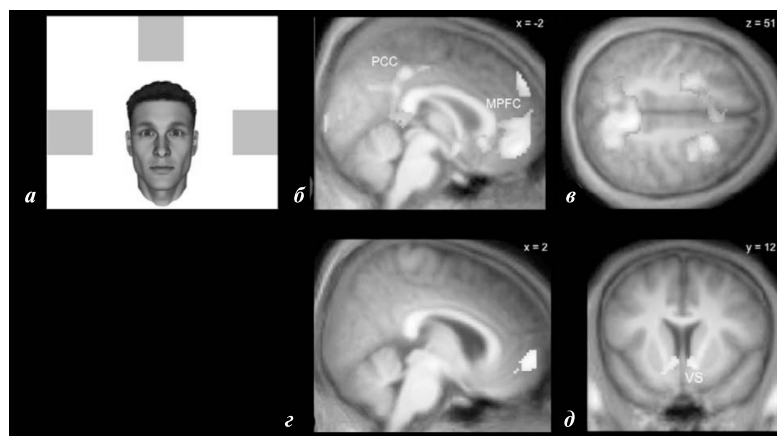
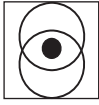


Рис. 2. Результаты нейровизуализации [Schilbach et al., 2010]:  
 а – экран, представляющий антропоморфного виртуального персонажа и три предмета (серые квадраты); б – нейронные корреляты совместного внимания; в – нейронные корреляты неразделяемого внимания; з – нейронные корреляты совместного внимания, инициированного другим; д – нейронные корреляты совместного внимания, инициированного самим испытуемым

Таким образом, при том, что испытуемый постоянно фиксирует взгляд на предмете на экране, имеют место существенные различия в нейронной активации в зависимости от того, направлен или нет взгляд испытуемого на объект вместе с другим. Зрительная фиксация предмета в координации с другим приводит к активации нейронной сети, связанной с восприятием ментальных состояний и коммуникативных намерений другого человека (а также с саморефлективным режимом по умолчанию). Ментальная система активировалась при простой координации взгляда испытуемого с взглядом (виртуального) другого без явного намерения оценивать его ментальные состояния. Таким образом, активация областей мозга, составляющих ментальную систему, возможна при совместном внимании без эксплицитного знания о ментальных состояниях других людей и, следовательно, эта система может играть роль в дорефлективном, практическом ноу-хау взаимодействия с другими.

Результат, полученный Шильбахом, имеет большое теоретическое значение: оказывается, нейронная сеть ментальной системы не связана только с рефлексивным, интеллектуальным социальным сознанием, но может служить коррелятом базовых механизмов, обеспечивающих первичные навыки социального взаимодействия. Возможно, именно такова первая стадия развития этой системы, и лишь позднее она начинает использоваться для явного, эксплицитного социального сознания.

В описываемом эксперименте при помощи МРТ были также выявлены различия нейронных коррелятов в зависимости от того, кем



была инициирована ситуация совместного внимания: когда испытуемый следовал взгляду другого, активировалась передняя зона медиальной предлобной коры, тогда как инициация совместного внимания испытуемыми приводила к активации вентральной области полосатого тела, входящей в систему, отвечающую за переработку информации о вознаграждении (подкреплении). Оценка субъективного опыта испытуемыми также показала, что инициируемая ими ситуация совместного внимания приносила удовольствие. Из этого можно сделать вывод о том, что социальное взаимодействие включает в себя аффилиативную мотивацию, реализация которой сама по себе служит вознаграждением, и, следовательно, взаимодействие и разделение опыта с другим – самостоятельный и важный мотив.

Мы представили два эксперимента, исходящих из новой экспериментальной парадигмы Шильбаха и др.; первый был связан с диадической ситуацией интересубъективности, а второй – с ситуацией триадической, ситуацией совместного внимания. Цель обоих экспериментов – создать более достоверную с экологической точки зрения ситуацию взаимодействия между субъектами. Оба были основаны на отслеживании взгляда, возможно, одном из самых характерных аспектов интересубъективного взаимодействия. Эти эксперименты подтверждают различие в нейронной активации между точками зрения наблюдателя и участника взаимодействия. Более того, оказывается, что базовое взаимодействие между людьми имеет дорефлективный, преинтеллектуальный характер – оно является практическим навыком, ноу-хау.

Эксперимент показывает, что новая философская аксиоматика, порывающая с интерналистской репрезентационистской аксиоматикой, которой руководствуются многие исследователи, может успешно руководить постановкой и интерпретацией экспериментов и в результате порождает более сложную, экологически достоверную и адекватную модель интересубъективности. «Новая» аксиоматика в действительности опирается на «старые» феноменологию и прагматизм, но сам факт возвращения к этой философии говорит о необходимости философии как таковой в интердисциплинарном научном пространстве, в частности в исследовании сознания. Всякий эксперимент интерпретируется исходя из определенной теории, на основе определенной аксиоматики – ведь многие эксперименты сами по себе теоретически недоопределены. Роль философского анализа – прояснить выбранную (зачастую подспудно, неосознанно) аксиоматику и предоставить рациональные основания для выбора одной из возможных теорий. Представляется, что прагматическая (не только в историческом, но и в этимологическом смысле) теория является лучшим кандидатом для руководства и прояснения экспериментальной деятельности, чем интернализм и репрезентационизм.



## Библиографический список

Станжевский Ф.А., Станжевский А.А. Некоторые методологические проблемы нейровизуализации функций социального сознания // Физиология человека (в печати).

Berthoz, Petit, 2006 – *Berthoz A., Petit J.L.* Phénoménologie et physiologie de l'action. P. : Odile Jacob, 2006.

De Jaegher, Di Paolo, 2007 – *Jaegher H. de, Di Paolo E.* Participatory Sense-Making. An Enactive Approach to Social Cognition. *Phenom. Cogn. Sci.*, 2007.

Gallagher, 2005 – *Gallagher Sh.* How the Body Shapes the Mind. N.Y. : Oxford University Press, 2005.

O'Regan, Noe, 2001 – *O'Regan J.K., Noe A.* A Sensorimotor Account of Vision and Visual Nonconsciousness // *Behavioral and Brain Sciences*. 2001. № 24.

Tronick, 1998 – *Tronick E.* Dyadically Expanded States of Consciousness // *Infant Mental Health Journal*. 1998. Vol. 19(3). P. 290–299.

Schilbach, et al., 2010 – *Schilbach L., Wilms M., Eickhoff S.B.* [et al.]. Minds Made for Sharing: Initiating Joint Attention Recruits Reward-Related Neurocircuitry // *Journal of Cognitive Neuroscience*. 2010. № 22(12).

Schilbach, et al., 2006 – *Schilbach L., Wohlschlaeger A.M., Kraemer N.C.* [et al.]. Being with Virtual Others: Neural Correlates of Social Interaction // *Neuropsychologia*. 2006. № 44(5).

Schilbach et al., in press – *Schilbach L., Timmermans B., Reddy V.* [et al.]. Toward a Second-Person Neuroscience // *Behavioral and Brain Sciences* (in press).

Wilms et al., 2010 – *Wilms M., Schilbach L., Pfeiffer U.* [et al.]. It's in your eyes-using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience // *Social Cognitive and Affective Neuroscience*. 2010. № 5(1).



## MIND AND BRAIN STATES: EMBEDDING THE MENTAL IN THE LIVING ORGANISM

**Inês Hipólito** – Centre  
for Artificial  
Intelligence, New  
University of Lisbon.  
E-mail: hipolito.ines  
@gmail.com



With neurons emergence, life alters itself in a remarkable way. This embodied neurons become carriers of signals, and processing devices: it begins an inexorable progression of functional complexity, from increasingly drawn behaviors to the mind and eventually to consciousness [Damasio, 2010]. In which moment has awareness arisen in the history of life? The emergence of human consciousness is associated with evolutionary developments in brain, behavior and mind, which ultimately lead to the creation of culture, a radical novelty in natural history.

It is in this context of biological evolution of conscious brains that we raise the question: how conscious brains connect with each other? In order to answer it, I will explore how brain states and conscious states each participate in dynamic interactive processes involving the whole organism. I will argue that a possible way to overcome the hard problem of consciousness might be based on the notion of embodiment as a process of embedding the mental in the living organism relating dynamically with the environment through the sensory-motor experience. In order to do so, I will provide an assembly between an anthropologic perspective of consciousness with contemporary Philosophy of Mind, Interaction Theory [Gallagher 2001, 2008; Zahavi 2001, 2008; Fuchs and De Jaegher 2009].

**Key words:** biological evolution; consciousness; interaction theory.

### 1. Consciousness on Biological Evolution

In the evolutionary process, how did the consciousness arise in a world where, until then, it did not exist? In animals' development, when did they first experienced mental events emerging from the darkness of unconsciousness?

These mental events must be related to the neuronal events that continually were posted in the brain, but how did they improved brain performance, making it so useful for evolutionary survival? In fact, the question that needs to be addressed is the following: what ontological status can be attributed to mental events that appeared in a world that could be once considered as a monistic physical world, this is, the world of matter and energy?

The instinctive performance of an animal is based on ontogenetic construction of the nervous system and related structures via genetic instructions. From the materialistic viewpoint, learning, meanwhile, can be regarded as an increasing the synaptic efficiency in repeated use. We can thus consider: which advantage has been produced by experiences associated with mental actions? William James suggested that the mind was a property acquired by a brain that is now too complex to control himself. It has been suggested that consciousness is useful regarding that it provided the animal a certain holistic experience. The immense diversity of standardized neuron activity conveyed coded information that could be used to reconstitute the image,



but a holistic operation like this cannot be performed by cortex machinery; it is, rather, achieved in conscious experience that arise when we open our eyes and that changes itself every moment in an apparent synchrony with visual information [see Hipólito, 2014a].

The Darwinian theory of evolution is imperfect since it does not recognize the extraordinary problem that is posed by the living organisms that acquire mental experiences of non-material type. Moreover and for another hand, the Cartesian solution cannot be accepted, given that it argues that humans conscious experiences are attributable to divine creation of souls, and therefore humans are mechanical automates empty of mental activity; as Popper [1982: 150] states “the emergence of consciousness in the animal kingdom is perhaps as great mystery as the origin of life itself. We must admit, however, and despite its impenetrable difficulty, that it is a product of evolution, of natural selection”.

### 1.1. Mind-Brain: the rising of Self-Awareness

The materialist critics argue that there are insuperable difficulties in the case of immaterial mental events such as thinking acting over material structures such as the neurons of the cerebral cortex. Such action is alleged to be inconsistent with the conservation laws of physics, in particular the first law of thermodynamics. This objection would, obviously, be supported by the physicists of the nineteenth century and by neuroscientists and philosophers who lie still ideologically in nineteenth-century physics, not recognizing the evolution produced by quantum physicists of the twentieth century. Unfortunately, it is rare for a quantum physicist, to dare meddling in the mind-brain issue. Margenau (1984), affords a fundamental contribution «that some fields, such as the field of mechanical probability quantum, do not transport energy or matter” which means a remarkable transformation of nineteenth-century physics. Margenau continues stating:

“In very complicated physical systems as the brain, neurons and the sense organs, whose constituents are small enough to be governed by quantum laws of probability, the physical body is always positioned for a number of possible modifications, each with a probability defined; if it occurs a modification which requires more energy or less energy than another, the intricate body provides it automatically. The mind would not be called upon to provide energy.” (p. 96).

Shortly, what Margeneau asserts is that: “mind can be seen as a field in the physical sense of the term. But, as a non material field, for which the closest analogy is, perhaps, a field of probability” (p. 97). According to Margeneau (1984), the hypothesis is that of the mind-brain intersection being analogous to a quantum mechanics probabilistic field, which has no



mass nor energy but that can, however, lead to effective action in microsites. It is proposed more specifically, that the intentions involved in mental concentration or planned thought may lead to neuronal events by a procedure analogous to quantum mechanics probability fields. In this context, one may question which neural events could be proper recipients of mental events, which are analogous of quantum probability fields. The answer can be found in recent discoveries about the nature of the synaptic mechanism by which a nerve cell communicates with another, specially regarding mini-sites operation. To this extent, the microsites in the brain may have transcendental capabilities to be channels of communication between these two distinctive entities (mind-brain). If one accepts the fact that mental events can effectively act on brain, then the philosophical implications are far reaching.

It is in this context of biological evolution of conscious brains that we raise the question: how conscious brains engage with each other? In order to answer it, we will explore how brain states and conscious states each participate in dynamic interactive processes involving the whole organism and therefore, that human brains learn by interaction and that this occurs at the level of the *second-person* perspective [Gallagher, 2001, 2008; Zahavi 2001, 2005; Reddy 2003; Fuchs, 2005]. In the early years of social neuroscience, attempts had already been made to investigate two brains in interaction through “hyper-scanning,” which was hailed as a break-through technology [Montague et al., 2002]. Although the application of this method has, indeed, provided invaluable insights into the neural basis of social cognition (SoCog) in conditions of health and pathology [e.g., King-Casas et al., 2005; 2008], the approach never really caught on. At least in part this is due to the fact that using it to its full potential would have required establishing more ecologically valid ways for two or more participants to interact [cf. Redcay et al., 2010]. However, first steps are now being taken to investigate the neural mechanisms of interaction dynamics.

Brain is an organ of modulation and transformation that mediates the cycles of organism-environment interaction. According to Fuchs (2009), the mind is not in the brain, but it is rather distributed among the brain, the body and the environment, this is, the brain is not only the organ of the mind since it is connected to a human head and body, and this body is connected to its environment and to other embodied human beings.

Contemporary Philosophy of Mind is mainly based on the assumption of a profound distinction between consciousness and biological life, in other words, the one conceived as internal and purely mental and the other as an external, functional property of physical systems: the so called hard problem of consciousness, which cannot be solved as long as the mind and life are conceptualized as exclusive concepts. However, according to new studies in Embodied Cognitive Science, the person must be thought not as pure subjectivity experienced from within and not as a complex physiolo-





gical system observed from without, but as a living being interacting with others within the second person perspective on the “you” perspective.

## 2. Brain, Body and World: Learning by Interaction

In Philosophy of Mind there is a duality between the first and third person perspective, but this seems insufficient since there is a gap between a person’s mental states and another’s perception of that person’s body which is the only experienced from within and the other observable from without.

To overcome the antagonism of first and third person perspective, it has recently been proposed to introduce the notion of the second person perspective [Gallagher, 2001, 2008; Zahavi 2001, 2005; Reddy 2003, Fuchs, 2005].

In spite of the remarkable progress made in the burgeoning field of neuroscience, the neural mechanisms that underlie social encounters are only beginning to be studied and could – paradoxically – be seen as “dark matter” [Schilbach et al., 2013]. The field of Neuroscience has begun to illuminate the complex biological bases of human social cognitive (SoCog) abilities [Frith & Frith, 2010; Ochsner & Lieberman, 2001]. Two new neuroanatomically distinct large-scale networks have gained center stage as the neural substrates of social cognition (SoCog): the so-called *mirror system* (MNS) and the *mentalizing network* (MENT). The former is believed to give us a “first-person grasp” of the motor goals and intentions of other individuals [Rizzolatti & Sinigaglia, 2010]. The latter has been seen as providing evidence for a “Theory-Theory” account of SoCog believed to give us an inferential, reflective and what might be called a *third-person* grasp of the others’ mental states (Frith & Frith, 2006, 2010). However it remains unclear whether, and how, activity in the large-scale neural networks described above is modulated by the degree to which a person does or does not feel actively involved in an ongoing interaction.

The duality of the first and third person perspective (1PP and 3PP) is an established opposition in philosophy of mind where it is mainly used to demonstrate the irreducibility of subject as against a physicality concept of the world [Fuchs, 2012]. The experiential perspective of a subject and the observational perspective (i.e. neuroscientist) cannot be brought to a final congruence because even the sum of any possible knowledge about objective process occurring in the subject’s brain, body and surrounding world would not include what it is like for the subject to have the experience in question [Nagel, 1974]. It has recently been proposed to introduce the notion of the second person or intersubjective perspective (2PP) in order to overcome the antagonism of first and third person or subjective and objec-



tive perspective [Gallagher, 2001, 2008; Zahavi, 2001, 2005: 206–214; Reddy, 2003; Fuchs, 2012].

From this context results a triad of perspectives that has gained importance in contemporary Philosophy of Mind and SoCog. Each theory defining the access we use in understanding other persons:

1. Theory of Mind or Theory-Theory, claims traditionally, that other minds are known by reference to the best suitable hypothesis on the reasons and motives for their behavior, this is, on the basis of observation, which means from a *third-person perspective*. This means that interacting with others does not add anything to this access in principle [Perner, 1991].
2. According to Simulation Theory, other minds are known by reference to a first person perspective. This means that understanding others requires to run an inner simulation of their behavior, thus creating an “as-if” mental state which then has to be somehow projected onto the other [Gallese and Goldman, 1998; Goldman, 2006].
3. The Interaction Theory is the most recent approach to Philosophy of Mind and SoCog and it claims that it is through immediate perception of, and embodied interaction with others that we gain our primary experience of their feelings and intentions, without recourse to inner theories or simulations. This approach focuses on the expressive bodily behavior, inter-bodily resonance, intentions as visible in action and the shared situational context in order to explain social understanding [Gallagher, 2001, 2008; Zahavi, 2001, 2008; de Jaegher and Di Paolo, 2007; Fuchs and De Jaegher, 2009].

The second-person approach has already begun to prove productive within SoCog, pointing out the importance of experiencing and interacting with others as our primary ways of knowing them, preliminary evidence from neuroimaging demonstrates profound differences in neural processing related to the reciprocity of social interaction [Schilbach et al., 2013], which is consistent with our proposal that the second-person perspective can make an important contribution to the neuroscientific study of social encounters and could, in fact, lead to the development of second person neuroscience. In respect to developing awareness of minds through second-person engagements the role of social interaction for cognitive and social development has begun to gain center in several scientific discourses [de Jaegher et al., 2010]: the role of interaction as a vehicle for the acquisition of knowledge has, for instance, been demonstrated in language development. In contrast with the Chomskian idea of a “language acquisition device” [Chomsky, 1979], the perception of structure in social engagement has been shown to guide vocal development and language learning [Bruner, 1983; Goldstein & Schwade, 2010], both in terms of speech perception or turn-taking.

In other areas of research interaction has been investigated by focusing on processes such as involuntary mimicry [Chartrand & Bargh, 1999; Nie-



denthal et al., 2010], which leads to enhanced rapport and linking, but is also influenced by differences in affiliate motives and independent self-construal [van Baaren et al., 2003; Baaren et al., 2004]. Social interaction, however, involves more complex forms of coordination present from early on in life [cf. Harrist & Waugh, 2002]. In this respect, Knoblich & Sebanz (2008) distinguish between “action simulation” [Rizzolatti & Sinigaglia, 2010], “joint attention” and “shared intentionality” [Tomasello & Carpenter, 2007]. The development of a shared perceptual and attentional space within early mutual attentional engagements is curtailed, because it paves the way for triadic interactions in which we share our mental states about a third object or person with others [Tomasello, 1995]. Such triadic attentional engagements may be crucial for developing complex joint actions [Fiebich & Gallagher, 2012].

A second-person account does not replace first and third person accounts, but bookends them to a large degree. Developing approaches for data collection and analysis from two interacting persons (and possibly two brains) is relevant, as is revisiting and modifying established experimental paradigms to incorporate an emotionally engaged, interactive perspective. How is it that largely automatic, implicit forms of interaction develop into explicit social cognitive capacities and how does explicit mind knowledge contribute to ongoing interactions? If one considers interaction as developmentally prior to mentalizing or mirroring, then the question remains: what drives interaction? How insights from a second-person account could be put to use in future research using computational neuroscience techniques and in the emerging field of social neuroendocrinology?

In this article, we argue for the conception of a second-person approach to other minds, suggesting that interpersonal understanding is primarily a matter of social interaction and emotional engagement with others. This hypothesis – well grounded in contemporary philosophical considerations – provides a different route to the investigation of intersubjectivity by emphasizing aspects of SoCog specifically related to the procedural nature and experiential aspects of social integrations [See Hipólito, 2014b].

## Conclusions

Although neurons have the capacity to send electrochemically signals, send that signals to a set of destinations in the body and generate circuits and systems of an enormous complexity, they are, nevertheless, body cells, profoundly reliant on nutrients – as all cells in the body – and that are distinguishable by their ability to do things that other cells cannot do.

The traditional Cartesian and ontological division between mind and brain is, to our knowledge, inconceivable, considering that the neurons that compose the brain are body cells and this fact must be considered closely



when we assess the hard problem of consciousness. As neurons emerge in the body interior, endowed with movement possibility, life revises itself in the way denied to plants. It begins an inexorable progression of functional complexity, from behaviors increasingly elaborated to mind itself, and, eventually, to consciousness.

The functional meaning of the number of neurons and their organizational patterns is the reason by which it is not possible to assess behavior and mind problems by excessively an isolated research on neurons or on molecules that act upon them, or the genes involved in the management of its life. Nervous systems develop themselves as life managers and healers of the biological value, at the beginning supported by innate provisions, but afterwards by images, this is, by minds. On biological evolution, behavior based in mind became very complex in numerous non-human species, however it is possible that flexibility and creativity that distinguish human performance could not appear in a genetic mind. [Damasio, 2010] Mind must have been carried and enhanced by the self-process that would come up in its core.

As the self comes to mind, life alters itself: interior and exterior mental images are now coherently organized by the proto-self [Damasio, 2010] and are now guided by the homeostatic demands of the body. SoCog begins now to be flexible, this is, the presence of the core self is followed by an expansion in cognitive processing space, of conventional memory, working memory and reasoning. Autobiographical self emerges and with this phenomenon, vital regulation changes radically. The emergence of consciousness is associated to evolutionary developments in the brain, behavior and mind that end conducting to culture, a radical novelty on the natural history route. The emergence of neurons with its consequent behavior diversification and preparation of the tack to the mind must be a remarkable even on a grand journey. However, the emergence of conscious brains capable of self-awareness is the grand event that follows.

A possible way to overcome the hard problem of consciousness might be based on the notion of embodiment as a process of embedding the mental in the living organism relating dynamically with the environment through the sensory-motor experience. Therefore, the brain must be considered primarily as an organ of the living being, and only by this becomes an organ of the mind. In such conditions, consciousness is not an object or state that can be localized: but a process relating to something [Fuchs, 2009]. It is possible to overcome the hard problem if mind, consciousness and life relate in co-existence from the second person perspective.

Lived embodiment on the one hand, and the physical body (including the brain) on the other hand (Leib and Körper) are two aspects of one living organism in relation to others – one corresponding to the first and the second person perspective, the other to the third person perspective. Having this considerations accepted, we might be able to overcome the two radically different ontology (the mental and the physical). In order to do so, we need to challenge the Cartesian idea that others' mental states are hidden



away and inaccessible and reject the notion that we ordinarily act as a spectator of others' behavior. Theory-Theory fails to grasp the primary way in which we relate to and interact with others. Learning by interaction or the second perspective approach explains how two minds connect with each other in a form of embodied practice that is emotion, sensory-motor, perceptual and non-perceptual. Second-person interactions comprehends the innate developing capacity to interact with others manifested at a perceptual experience and, at this level, the others' mind is not entirely hidden or private, but is given and manifested in others' embodied behavior.

## References

- Bruner, J. (1983) *Child's talk*. Norton, 1983.
- Chartrand, T.L. & Bargh, J.A. (1999) *The chameleon effect: The perception – behavior link and social interaction* // Journal of Personality and Social Psychology. 1999. Vol. 76(6):893–910. Available at: <http://dx.doi.org/10.1037/0022-3514.76.6.893>.
- Chomsky N. (1979) *Language and Responsibility*. Harvester Press.
- Cooper J.C., Dunne S., Furey T. & O'Doherty JP. (2012) *Human dorsal striatum encodes prediction errors during observational learning of instrumental actions* // Journal of Cognitive Neuroscience. 2012. № 24(1). P. 106–118. Available at: [http://dx.doi.org/10.1162/jocn\\_a\\_00114](http://dx.doi.org/10.1162/jocn_a_00114).
- Creutzfeldt, O.D. (1979). *Neurophysiological Mechanisms and consciousness*. *Ciba Found Symp* 69. P. 217–233.
- Creutzfeldt, O. D. (1987). *Inevitable deadlocks of the Brain-Mind discussion*. In B. Guylas (ed.) *The Brain-Mind Problem*. Assen/Maastricht: Leuven University Press.
- Damasio, A. (2010). *Self comes to Mind*. New Iorque: Random House.
- De Jaegher, H. & Di Paolo, E. (2007) Participatory sensemaking. An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences* 6(4):485–507. Available at: <http://www.dx.doi.org/10.1007/s11097-007-9076-9>.
- Dobzhanski, T. (1967). *The Biology of Ultimate Concern*, New York: The New American Library
- Fiebich, A. & Gallagher, S. (2012) Joint attention in joint action. *Philosophical Psychology*. DOI:10.1080/09515089.2012.690176. Published online May 28, 2012.
- Frith, C. D. & Frith, U. (2008) Implicit and explicit processes in social cognition. *Neuron* 60(3):503–10. Available at: <http://dx.doi.org/10.1016/j.neuron.2008.10.032>.
- Frith, U. & Frith, C. D. (2010) The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365(1537):165–76. Available at: <http://dx.doi.org/10.1098/rstb.2009.0160>.
- Froese, T. & Fuchs, T. (2012) The extended body: A case study in the neurophenomenology of social interaction. *Phenomenology and the Cognitive Sciences* 11(2):205–35.
- Fuchs, T. (2009) *Das Gehirn – ein Beziehungsorgan. Eine phänomenologisch-ökologische Konzeption*. Kohlhammer.



- Gallagher, S. (2001) The practice of mind: Theory, simulation, or interaction? *Journal of Consciousness Studies* 8(5–7):83–108. Available at: <http://pegasus.cc.ucf.edu/~gallagher/practice01.htm>.
- Gallagher, S. (2005) *How the body shapes the mind*. Oxford University Press.
- Gallagher, S. (2008) Direct perception in the intersubjective context. *Consciousness and Cognition* 17(2):535–43. Available at: <http://dx.doi.org/10.1016/j.concog.2008.03.003>.
- Gallese, V. & Goldman, A. (1998) Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences* 2:493–501.
- Goldman, A. (2006) *Simulating minds. The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press
- Goldstein, M. H. & Schwade, J. (2010) From birds to words: Perception of structure in social interactions guides vocal development and language learning. In: *The Oxford handbook of developmental behavioral neuroscience*, ed. M. S. Blumberg, J. H. Freeman & S. R. Robinson, pp. 708–29. Oxford University Press.
- Grossmann, T. & Johnson, M. H. (2010) Selective prefrontal cortex responses to joint attention in early infancy. *Biology Letters* 6(4):540–43. Available at: <http://dx.doi.org/10.1098/rsbl.2009.1069>
- Hipólito, I. (2014a) Multisensory Integration: A Neurophilosophical account of Biology of Perception and Loss of Self-evidence in disorders of disintegration. *Analytical and Continental Philosophy: Methods and Perspectives*. Austria: De Gruyter
- Hipólito, I. (2014b). «On Autism and Interaction Theories of the Self». In Gerner, A. Gonçalves, J. (eds.) *Altered Self and Altered Self-Experience*, Norderstedt: BoD, 204 (296 pp.) ISBN 978–3-7357–6004–3
- Hull, C. L. (1943) *Principles of behavior*. Appleton-Century-Crofts.
- Hutto, D. D. (2008) *Folk-psychological narratives. The sociocultural basis of understanding reasons*. MIT Press.
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P. & Montague, P. R. (2008) *The rupture and repair of cooperation in borderline personality disorder*. *Science* 321(5890):806–10. Available at: <http://dx.doi.org/10.1126/science.1156902>.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R. & Montague, P. R. (2005) Getting to know you: Reputation and trust in a two-person economic exchange. *Science* 308(5718):78–83. Available at: <http://dx.doi.org/10.1126/science.1108062>.
- Knoblich, G. & Sebanz, N. (2008) Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society London B: Biological Sciences* 363(1499):2021–31. Available at: <http://dx.doi.org/10.1098/rstb.2008.0006>.
- Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., Apple, N. & Fischer, R. E. (2002) Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage* 16(4):1159–64. Available at: <http://dx.doi.org/10.1006/nimg.2002.1150>.
- Nagel, T. (1974) «What Is it Like to Be a Bat?», *Philosophical Review*, pp. 435–50.
- Niedenthal, P. M., Mermillod, M., Maringer, M. & Hess, U. (2010) The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences* 33(6):417–33; discussion 433–80. Available at: <http://dx.doi.org/10.1017/S0140525X10000865>.



- Noë, A. (2009) *Out of our heads*. Hill & Wang.
- Ochsner, K. N. & Lieberman, M. D. (2001) The emergence of social cognitive neuroscience. *American Psychologist* 56(9):717–34. Available at: <http://dx.doi.org/10.1037/0003-066X.56.9.717>.
- Popper, K. (1982). *The Open Universe: An Argument for Indeterminism*, London: Hutchinson
- Popper, K. Eccles, J. (1977). *The Self and its Brain*. Berlin: Springer-verlag.
- Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. & Saxe, R. (2010) Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage* 50(4):1639–47. Available at: <http://dx.doi.org/10.1016/j.neuroimage.2010.01.052>.
- Reddy, V. (2008) How infants know minds. Harvard University Press.
- Rizzolatti, G. & Sinigaglia, C. (2010) The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience* 11 (4):264–74. Available at: <http://dx.doi.org/10.1038/nrn2805>.
- Schilbach, L. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences* 36, 393–462, doi:10.1017/S0140525X12000660
- Schilbach, L. (2010) A second-person approach to other minds. *Nature Reviews Neuroscience* 11(6):449. [arLS]
- Schilbach, L., Bzdok, D., Timmermans, B., Fox, P. T., Laird, A. R., Vogeley, K. & Eickhoff, S. B. (2012a) Minds at rest revisited: Using ALE meta-analyses to investigate commonalities in the neural correlates of socio-emotional processing and unconstrained cognition. *PLoS One*. 7(2): e30920.
- Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R. & Vogeley, K. (2006) Being with virtual others: Neural correlates of social interaction. *Neuropsychologia* 44(5):718–30. Available at: <http://dx.doi.org/10.1016/j.neuropsychologia.2005.07.017>.
- Tomasello, M. & Carpenter, M. (2005) The emergence of social cognition in three young chimpanzees. *Monograph of the Society for Research in Child Development* 70(1): vii–132.
- Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28:675–91.
- Van Baaren, R. B., Holland, R. W., Kawakami, K. & van Knippenberg, A. (2004) Mimicry and pro-social behavior. *Psychological Science* 15(1):71–77. Available at: <http://dx.doi.org/10.1111/j.0963-7214.2004.01501012.x>.
- Van Baaren, R. B., Maddux, W. W., Chartrand, T. L., De Bouter, C. & Van Knippenberg, A. (2003) It takes two to mimic: Behavioral consequences of self-construals. *Journal of Personality and Social Psychology* 84(5):1093–102. Available at: <http://dx.doi.org/10.1037/0022-3514.84.5.1093>.
- Weiskrantz, L. (1974). “The interaction between occipital and temporal cortex in vision: an overview”. In F. O. Schmitt & F. G. Worden (eds.) *The Neurosciences*. 3rd Study Program, Cambridge: MIT Press
- Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R. & Vogeley, K. (2010) It’s in your eyes – using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience* 5(1):98–107. Available at: <http://dx.doi.org/10.1093/scan/nsq024>.
- Zahavi, D. (2005) Subjectivity and selfhood: Investigating the first-person perspective. The MIT Press.



## BLOCKERS: A REPLY TO HAWTHORNE

**James Grindeland** –  
University of North  
Georgia. E-mail:  
james.grindeland  
@gmail.com



Physicalism is roughly the thesis that everything is physical. The two most popular ways of formulating physicalism rigorously are the ways given by Frank Jackson and David Chalmers. The best objections, in turn, include John Hawthorne's 'blocker' objections. Hawthorne argues that, if it is possible for there to be non-physical beings or properties that prevent certain mental phenomena from existing (i.e., non-physical blockers), Jackson's and Chalmers' formulations will be inadequate. Jackson's formulation will be inadequate by virtue of not capturing all of the right physical dependence principles. Chalmers' formulation will be inadequate in so far as, when modified to define 'restricted physicalisms', such as physicalism of the mental, the restricted formulations will not capture all of the right physical dependence principles. By contrast, I argue that Hawthorne's blocker arguments are misguided on the grounds that non-physical blockers are impossible; I argue that his critique of Chalmers' formulation is unsound by virtue of falsely presupposing that restricted physicalisms require restricted formulations of physicalism (I argue that it is only necessary to define physicalism of a world); and I argue that Jackson's and Chalmers' formulations capture all of the right physical dependence principles.

**Key words:** Blockers, Physicalism.

Physicalism is roughly the thesis that everything is physical. The two most popular ways of formulating physicalism rigorously are the ways given by Frank Jackson (1994, 1998) and David Chalmers (1996). The best objections, in turn, include John Hawthorne's (2002) 'blocker' objections. Hawthorne argues that, if it is possible for there to be non-physical beings or properties that prevent certain mental phenomena from existing (i.e., non-physical blockers), Jackson's formulation will be inadequate by virtue of not capturing all of the right physical dependence principles, and Chalmers' formulation will be inadequate in so far as, when modified to define 'restricted physicalisms' – such as *physicalism of the mental* – the restricted formulations will not capture all of the right physical dependence principles.

By contrast, I argue that Hawthorne's blocker arguments are misguided on the grounds that non-physical blockers are impossible; I argue that his critique of Chalmers' formulation is unsound by virtue of falsely presupposing that restricted physicalisms require restricted formulations of physicalism; and I argue that Jackson's and Chalmers' formulations capture all of the right physical dependence principles.

### Jackson's formulation

Jackson's formulation: Physicalism is true of our world iff "Any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world" [Jackson, 1994: 28; Jackson, 1998: 12]. Understanding this requires understanding the jargon: 'minimal physical duplicate', 'duplicate simpliciter', 'physical'.





First: for Jackson, *minimal physical duplicates* are physical duplicates in the sense of being “physical property and relation for physical property and relation identical” [Jackson, 1994: 27]. This requires identity of the physical *qualities*, but doesn’t require identity of the *individuals* that instantiate them. Physical duplicates need merely be *qualitatively* physically the same. Yet not all physical duplicates are *minimal* physical duplicates. Being a minimal physical duplicate of a given world *w* requires meeting the further condition of not having any ‘ingredients’ that are not required in order to be a physical duplicate of *w*. Jackson likens the qualification of ‘minimal’ to a ‘stop’ clause in food recipes. “Of necessity the writers of recipes rely on an intuitive understanding of an implicitly included ‘stop’ clause in their recipes” [Jackson, 1994: 28]; “otherwise they would face the impossible task of listing all the things you should *not* do” [Jackson, 1998: 12–13], such as “not to add dirt to the flour . . . don’t add bats wings” [Jackson, 1994: 28], etc. Thus: “A minimal physical duplicate of our world is what you would get if you – or God, as it is sometimes put – used the physical nature of our world (including of course its physical laws) as a recipe in this sense for making a world” [Jackson, 1994: 28].

Altogether, a minimal physical duplicate of a given world *w* would be a physical qualitative duplicate of *w* that doesn’t have any extra ‘ingredients’ that are not required in order to be a physical qualitative duplicate of *w*.<sup>1</sup>

Second: for Jackson, a given world is a *duplicate simpliciter* of *w* iff it is qualitatively identical to *w* in every way (physically and otherwise). Thus a duplicate simpliciter of *w* will have all of the same qualitative truths as *w*, and it will have no other positive qualitative truths – i.e., *that’s all* the positive qualitative truths it will have. Note: here there is another ‘that’s all’ or ‘stop’ clause. Thus there are two ‘stop’ clauses in Jackson’s formulation of physicalism: the first is implied by ‘minimal’; the second is implied by ‘simpliciter’.

Third: in Jackson’s 1982, 1986, 1994 papers, he makes clear that by ‘physical’ he has in mind truths from physics, chemistry, biology or neurophysiology, and anything that follows from these truths logically or causally. In those papers, he doesn’t take a stand on whether all physical truths follow from the *microphysics*; however, in his (2001) paper co-written with Chalmers, he does hold that all physical truths follow from the microphysics. So, if we apply Jackson’s 2001 commitments backwards to understand his earlier formulations of physicalism, we can read his earlier formulations as defining the physical in terms of the microphysical.

Altogether, we can read Jackson as defining ‘physicalism’ as follows: Physicalism is true of world *w* iff any minimal microphysical qualitative duplicate world would be a qualitative duplicate simpliciter of *w*. In other words: Physicalism is true iff, necessarily, having *w*’s positive microphysical qualitative truths and no other positive basic qualitative truths implies having *w*’s positive qualitative truths and no other positive qualitative truths:

<sup>1</sup> Henceforth I will tend to keep the qualification of ‘qualitative’ implicit.



$D_{jac}$  Physicalism is true of  $w$  iff  $\square [(w$ 's positive microphysical truths and *that's all* the positive basic truths)  $\supset$  ( $w$ 's positive truths and *that's all* the positive truths)]

Intuitively: Physicalism is true iff any world that has the same qualitative arrangement of basic ingredients (assuming they're microphysical) will have the same qualitative truths.

### Chalmers' formulation

"*Physicalism* . . . is true if . . . any world that is physically indiscernible from our world contains a copy of our world as a (proper or improper) part" [Chalmers, 1996: 41–42]. Here there are three implicit qualifications. First, by "physically" Chalmers means "microphysically." For he writes: "talk of physical properties is implicitly restricted to the class of fundamental properties unless otherwise indicated. I will sometimes speak of 'microphysical' or 'low-level physical' properties to be explicit" [Chalmers, 1996: 33]. Second, Chalmers means the 'if' to be read as an 'iff' so as to make it a definition, and not merely to indicate a sufficient condition. Third, by 'copy' Chalmers means qualitative copy; for, like Jackson, he is interested in qualitative properties, not the individuals or *haecceities* that have them.

Altogether, Chalmers defines 'physicalism' as follows: Physicalism is true of our world iff every microphysically qualitatively indiscernible world has a qualitative copy of our world as a part. Generalizing on our world: Physicalism is true of a given world  $w$  iff every world that is microphysically qualitatively indiscernible from  $w$  has a qualitative copy of  $w$  as a part, hence iff every world that is microphysically qualitatively indiscernible from  $w$  has *at least* the qualitative positive truths of  $w$ . That is:

$D_{cha}$  Physicalism is true of  $w$  iff  $\square [(w$ 's positive microphysical truths and *that's all* the positive microphysical truths)  $\supset$  ( $w$ 's positive truths)]

Intuitively: Physicalism is true iff any world that has the same qualitative arrangement of microphysical ingredients will have *at least* the same positive qualitative truths.

Note that, by contrast with  $D_{jac}$ ,  $D_{cha}$  only gives one 'stop' clause—and it is a different one. Namely, whereas  $D_{jac}$  says "*that's all* the positive basic truths" (microphysical or micro-nonphysical),  $D_{cha}$  says "*that's all* the positive microphysical truths." *Ceteris paribus*, this makes physicalism *prima facie* less likely to be true on  $D_{cha}$ ! For, on  $D_{cha}$ , physicalism's truth requires that our world's mental properties be instantiated in any microphysically indiscernible world, regardless of which basic nonphysical truths obtain. Hawthorne (2002) emphasizes this difference, as we will see next.



## Hawthorne's blocker arguments

Hawthorne (2002) gives 'blocker' arguments against  $D_{jac}$  and  $D_{cha}$ . Understanding them requires understanding what he means by a 'blocker'. A blocker would be any being or property<sup>2</sup> whose instantiation prevents ('blocks') the instantiation of certain mental properties. For instance, there might be some brain state  $S$ , a blocker  $B$ , and a mental state  $M$ , such that  $S \& \sim B$  entails  $M$ , but  $S \& B$  entails  $\sim M$ . In such a case,  $B$  is a blocker with respect to  $M$ .

Hawthorne is particularly interested in the possibility of *non-physical* blockers. A non-physical blocker would be a non-physical property whose instantiation prevents the instantiation of some mental property. Hawthorne doesn't profess to know whether non-physical blockers are possible; he is agnostic on this matter. Nevertheless, he argues that we should not define physicalism in such a way that *assumes* they are impossible; for, as he puts it: "when defining physicalism, one should not take a stand on controversial modal issues about which, intuitively, the materialist qua materialist has no commitments" [Hawthorne, 2002: 110].

Understanding Hawthorne's blocker arguments also requires grasping the physical dependence principle that he thinks any suitable definition of physicalism should capture. He keys us in on this physical dependence principle as follows: "Materialists about mentality will not tolerate the existence of mental facts, positive or negative, that have a robust explanation that appeals to immaterial entities" [Hawthorne, 2002: 109]. Given this, Hawthorne must endorse the following physical dependence principle:

PD Physicalism is true only if every truth (positive or negative) depends solely on physical truths (positive or negative), hence only if every (instantiated or non-instantiated) mental property depends solely on (instantiated or non-instantiated) physical properties.

According to PD, physicalism is false if any positive truths depend on the absence of any non-physical blockers. Thus physicalism is false if some positive mental truth obtains partly because some negative non-physical blocker truth obtains. That is, physicalism is false if some mental phenomenon depends on the absence of some non-physical blocker; physicalism would be false given this kind of *dependence on absence*. More precisely,

DA Physicalism is false if there is a positive mental truth  $M$  and a possible non-physical blocker  $B$  such that  $M$  obtains partly because  $B$  doesn't obtain.

DA gives a recipe for a counterexample to physicalism. In a counterexample of this sort, some positive truth would not depend solely on physical

<sup>2</sup> I say 'being or property', but for simplicity I will tend to speak of blockers as properties.



truths, but would depend at least partly on the non-existence of some non-physical blocker. (Hawthorne assumes the non-existence of a non-physical blocker would be a non-physical truth.)

Hawthorne judges that any suitable definition of physicalism should imply PD; so, since PD straightforwardly implies DA, he is committed to holding that physicalism should imply DA. Yet  $D_{jac}$  doesn't imply DA. For  $D_{jac}$ , taken as a 'physicalism algorithm' that 'checks' a given class of worlds for counterexamples to physicalism, doesn't 'check' the microphysically indiscernible worlds that might have non-physical blockers; recall: it only 'checks' the *minimal* microphysical duplicates. Thus, in so far as  $D_{jac}$  checks for necessities between various kinds of physicality and mentality, it only checks for necessities across certain worlds that lack non-physical blockers. So, as Hawthorne puts it:  $D_{jac}$  might be able to correctly assess whether a world's positive physical truths P "weakly necessitate" [Hawthorne, 2002: 103 and 111 fn. 1] its positive mental truths M in the sense that all of the P-worlds that don't have blockers are M-worlds; however, if non-physical blockers are possible,  $D_{jac}$  won't correctly assess whether P "metaphysically necessitate [s]" [Hawthorne, 2002: 103, 111 fn. 1] M in the sense that all the P-worlds (including those with non-physical blockers) are M-worlds.

Although  $D_{jac}$  does not imply DA,  $D_{cha}$  does imply DA. For  $D_{cha}$  does not require that any counterexample to physicalism's truth be a *minimal* microphysical duplicate of  $w$ ; it merely requires that it be *microphysically indiscernible* from  $w$  regardless of whether it is discernible from  $w$  in non-physical ways. Thus  $D_{cha}$  allows that worlds with non-physical blockers *may* function as counterexamples to physicalism's truth of  $w$ , as such worlds may be microphysically indiscernible from  $w$ . Hawthorne is happy with  $D_{cha}$  in this regard.

Yet Hawthorne nevertheless finds fault with  $D_{cha}$ . Namely,  $D_{cha}$ , like most definitions of physicalism, only formulates what it would take for physicalism to be true of a world, whereas many philosophers also want a formulation of what it would take for physicalism to be true of a 'proper part' of a world—say, the conscious beings. Thus he writes: "Philosophers are interested not merely in providing an account of physicalism *simpliciter*, but also an account of physicalism about a certain class of properties P (mental properties, for example) – call these 'restricted physicalisms'" [Hawthorne, 2002: 108].

Hawthorne's only gripe with  $D_{cha}$ , then, is that, if  $D_{cha}$  were to be suitably modified to say what it would take for physicalism to be true of some restricted class of properties—some 'proper part' of the world—the modified version would have a blocker problem. For the way Chalmers would modify  $D_{cha}$  to test for 'restricted physicalisms' is as follows: "It is clear how Chalmers thinks such a definition would go. Physicalism about, say, the mind, is true just in case every world physically indiscernible from the ac-



## BLOCKERS: A REPLY TO HAWTHORNE

tual worlds [sic] will have at least the positive mental facts that obtain at the actual world” [Hawthorne, 2002: 108].<sup>3</sup> That is:

$rD_{cha}$  Physicalism is true of the mental truths iff every physically indiscernible world has at least the positive mental facts that obtain in our world.

Moreover,  $rD_{cha}$  implies the following restricted version of PD:

$rPD$  Physicalism is true of the mental truths only if the mental truths (positive or negative) depend solely on physical truths (positive or negative), hence only if every (instantiated or non-instantiated) mental property depends solely on (instantiated or non-instantiated) physical properties.

Now Hawthorne finds fault with  $rD_{cha}$ . For suppose there is a non-physical blocker that blocks the instantiation of some mental truth. In that case,  $rPD$  will rightly judge physicalism to be false of the mental, since there is a negative mental truth that doesn't depend solely on positive or negative physical truths, but depends partly on the existence of the non-physical blocker. Yet, as we will see,  $rD_{cha}$  is not guaranteed to yield this same correct result.

To give an example of a case where  $rPD$  yields the correct result but  $rD_{cha}$  yields the incorrect result, suppose a given world  $w$  has some positive mental truths. Now consider the set of worlds that are microphysically indiscernible from  $w$ : say,  $\{w, w_1\}$ . Suppose that, of these worlds,  $w$  has exactly one non-physical blocker,  $B$ , and  $B$  blocks some mental property, whereas  $w_1$  doesn't have any non-physical blockers. That is, suppose the distribution is as follows:

$w$ 's non-physical blockers:  $B$

$w_1$ 's non-physical blockers: none

Is physicalism true of  $w$ 's mental properties?  $rPD$  and  $rD_{cha}$  disagree.  $rPD$  judges physicalism to be **false** of  $w$ 's mental properties, since  $w$  has a negative mental truth that depends partly on the existence of its non-physical blocker,  $B$ . Yet  $rD_{cha}$  judges physicalism to be **true** of  $w$ 's mental properties, since all of the worlds that are physically indiscernible from  $w$  (i.e.,  $w$  and  $w_1$ ) have *at least* the mental properties of  $w$ . So, given the possibility of non-physical blockers,  $rD_{cha}$  and  $rPD$  can yield contradictory results. Hawthorne takes this as a reason to reject  $rD_{cha}$ . Indeed, he goes further and takes this problem for  $rD_{cha}$  as an indirect problem for  $D_{cha}$  on the grounds that  $rD_{cha}$  is how  $D_{cha}$  would be 'suitably modified' to define physicalism of the mental.

<sup>3</sup> Analogous general schema: Physicalism is true of the  $X$  truths iff every physically indiscernible world has at least the positive  $X$  truths that obtain in our world.



In sum, there are two non-physical blocker cases Hawthorne is worried about. The first is a case where a positive mental truth obtains partly because of the non-existence of some non-physical blocker. This is the option that threatens  $D_{jac}$ . In such a case, PD implies that physicalism is false, since there is an instantiated mental property that depends partly on the non-instantiation of some non-physical blocker. The second is a case where a negative mental truth obtains partly because of the instantiation of a non-physical blocker. This is the option that threatens  $\neg D_{cha}$ . In such a case,  $\neg$ PD implies that physicalism is false, since there is a non-instantiated mental property whose non-instantiation depends partly on the instantiation of some non-physical blocker.

According to Stoljar's (2009) *Stanford Encyclopedia of Philosophy* entry for 'Physicalism', "[Hawthorne's] blockers problem... is an open question in the literature."

## Objections to Hawthorne's blocker arguments

I argue that Hawthorne's blocker arguments are misguided, and that the blocker argument that is supposed to threaten  $D_{cha}$  is unsound.

Hawthorne's blocker arguments are misguided. For non-physical blockers are incoherent, hence impossible, and their impossibility renders Hawthorne's blocker arguments misguided on the grounds that they criticize definitions of physicalism for not accommodating the possibility of things (non-physical blockers) that aren't even possible in the first place.

My argument that non-physical blockers are impossible requires that they be incoherent. Their incoherence is partly due to the nature of physics. I will not claim to give *the* definition of physics; I will simply stipulate a definition that is useful for my purposes. And I am not going to define physics as the fundamental science of everything; for on that definition physicalism would be trivially true. I want a definition of physics according to which the question of physicalism is non-trivial. Thus I will follow Chalmers' practice of avoiding this trivialization by defining physics as "the fundamental science developed to explain observations in the *external world*" [Chalmers, 1996: 118–119, italics added].<sup>4</sup> More precisely, I will define science as the science of the fundamental beings and processes in the external world. By defining physics as being about the external world, this leaves open the possibility that science might not adequately study the 'internal world' of conscious experience. This way, we can imagine physicalism being false: physicalism will be false if our world's conscious expe-

<sup>4</sup> Accordingly, we should read  $D_{jac}$  and  $D_{cha}$  as implicitly speaking, not simply of microphysical truths, but of microphysical truths in the external world; likewise, we should read  $D_{jac}$  as speaking, not simply of basic truths, but of basic truths in the external world.



riences are not entailed by the external world's fundamental beings and processes.

Given this definition of 'physics', consider an arbitrary blocker B. If B is a fundamental being or process in the external world, B will be physical and therefore won't be a non-physical blocker. Likewise, if B is a non-fundamental being in the external world but is entailed by the fundamental beings and processes in the external world (which are themselves physical on this definition of 'physics'), B will again be physical. By contrast, there are only two ways that B might be non-physical.

First, B might be a non-fundamental being in the external world that isn't entailed by the external world's fundamental beings and processes: in such a case B would be said to *strongly emerge* from those fundamental beings and processes. This is an incoherent option because of the kind of bruteness that it would involve. Not all bruteness is bad: presumably every possible world has a brute micro state—brute facts about which fundamental beings and processes exist and how they're arranged. We can call this a bruteness of the micro-world. Strong emergence, however, would be a bruteness of what *follows* from the micro-world. It would be a kind of bruteness where the micro-world *makes* the macro-world happen, but without entailing it. This is perhaps not flatly contradictory, as in theory we can at least entertain proposals such as that there are nomologically necessary connections that are not metaphysically necessary. Maneuvers like this, however, require assigning causal work to laws. The idea would be that different possible laws can serve to make different possible macro-worlds supervene on the same base micro-world. On this option, laws would be blockers. It is, however, incoherent to assign causal work to laws, as laws are descriptions of patterns, not reasons for patterns. Or, perhaps better: even if we were to assign causal work to the laws that bridge external micro-worlds with their corresponding external macro-worlds, the laws would nevertheless have to be part of the external micro-world in order to do that, in which case they would be trivially entailed by the external world's fundamental beings and processes, in which case they would be physical blockers, contrary to hypothesis. So, the option that B is a non-fundamental being in the external world that isn't entailed by the external world's fundamental beings and processes is inconsistent, as it requires either that impotent laws are potent, or that potent laws are both entailed and not entailed by the external world's fundamental beings and processes. Either way there is inconsistency; so this option is impossible.

Second, B might be no part of the external world. This would require that B only be part of an internal world; B would be entirely mental. Since the other option for how B might be non-physical is impossible, this is the only way left that B might be non-physical. But could a purely mental being or process function as a blocker? On the face of it, one might hold that ghosts or gods could be entirely mental blockers. Upon deeper reflection, however, it seems that neither ghosts nor gods could be entirely men-



tal. To be entirely mental, they would have to be absent in the external world—completely absent, not simply hidden relative to some epistemic agents. Yet ghosts are traditionally imagined as floating around, manipulating billiard balls when nobody is looking, haunting houses, and doing other things in the external world: this requires that ghosts be physical by our definition of ‘physics’. It doesn’t matter that ghost hypotheses are not currently postulated by scientists; for the *true physics* doesn’t have to agree with our educated guesses as to the true physics. Nor would it matter if ghost hypotheses are not falsifiable, as that wouldn’t preclude ghosts from being postulated by the true physics; at best, it would preclude ghosts from being postulated by responsible scientists. All that matters, for our purposes is that ghosts are in the external world and are therefore physical. If they are fundamental, the true physics postulates their existence. If they are non-fundamental, the fundamental beings and processes entail them. Either way, they are physical, in which case either way they are not *non-physical* blockers, contrary to hypothesis. So, the option that there are ghosts who serve as non-physical blockers is inconsistent and therefore impossible.

Gods, too, may be considered part of the external world. Gods have been conceived as residing in outer space, watching us from above and hurling down rain and lightning bolts. Likewise, some gods have been conceived as hiding in dimensions that mainstream physicists haven’t been able to find; they, too, would count as residing in the external world—namely, in its hidden folds. What, then, would it take for a god to not be in the external world? At a minimum, this would require the following: such a god either (a) *resides in* a conscious experience, or (b) *literally is* a conscious experience. Either way, the conscious experience must not be in the external world or be entailed by anything in the external world; otherwise it would be physical. Thus, for a god to be a non-physical blocker, it would have to either be in, or literally be, a conscious experience. The same would be true of any non-physical blocker.

More generally, then, for *anything* to be a non-physical blocker, it would have to either reside in, or else literally be, a conscious experience. Yet it is difficult to make sense of how such a being might do its job of blocking mental properties. Recall the way blocking would work: there would be some physical property *P*, some blocker *B*, and some mental property *M*, such that  $P \& \sim B$  entails *M*, but  $P \& B$  entails  $\sim M$ . Yet we have concluded that non-physical blockers must either reside in, or literally be, conscious experiences. Thus, for there to be a non-physical blocker, there must be some physical property *P*, some conscious experience *C*, some (proper or improper) part *c* of *C*, and some mental property *M* such that  $P \& \sim c$  entails *M*, but  $P \& c$  entails  $\sim M$ . In such a case, *c* serves as a blocker with respect to *M*. How might *c* do its blocking work? One option is that *c* is a conscious experience that is roughly a ‘blob’ that goes around interfering with other conscious (or otherwise conscious) beings or processes. For instance, Fred might be thinking about pancakes and all of a sudden he starts thinking





## BLOCKERS: A REPLY TO HAWTHORNE

about roller coasters due to the interference of a conscious experience that came ‘crashing in’ or that in some way interfered with his conscious experience. This option is not totally absurd on the face of it; however, the ‘blobbiness’ and interference implies that it would be part of an arena, hence part of the external world, which would render it physical. Thus *c* could not be a *non-physical* blocker. The alternative option is that *c* is a conscious experience, or something that resides in a conscious experience, but where *c*’s blockings happen in a way that doesn’t require that *c* or *c*’s conscious experience be physical. Here there are two options: (i) the blocking happens by magic, or (ii) the blockings consist simply in the trivial logical consequence that *c* precludes its own negation. Yet magic is incoherent; so we can dismiss that option as impossible. Meanwhile, the latter kind of blocking work—that *c* precludes its own negation—is irrelevant, as such a blocker would not thereby prevent a *positive* mental property *M* from arising from a physical property *P*; the only blocking work *c* could do would be to preclude  $\sim c$ , which would be to block a negative truth. Thus any such blocker would fail to block a positive mental truth. This kind of blocker is therefore not the kind of blocker we’re looking for; it’s not the kind of blocker that might, as Hawthorne puts it, serve as a “wedge”<sup>5</sup> between the physical and the mental.

Altogether, we need not worry about the possibility of *non-physical blockers* when formulating physicalism. For any such blocker, whether it be a ghost, a god, or anything else, it would either be part of the external world, in which case it would be physical (hence not non-physical, contrary to hypothesis), or else it would not be part of the external world, in which case its blocking work would either be magical (hence incoherent) or logically trivial (hence irrelevant). The first option is inconsistent: such blockers would be physical and non-physical. The second option is incoherent or irrelevant. The only option that is neither inconsistent nor irrelevant is incoherent: this is the option that non-physical blockers are parts of conscious experiences and do their blocking work by magic. I certainly don’t want to be rash in assuming that magic’s incoherence renders it impossible. Perhaps something that seems incoherent to me can be possible after all. Indeed, I am sympathetic with Hawthorne’s point that “when defining physicalism, one should not take a stand on controversial modal issues about which, intuitively, the materialist qua materialist has no commitments” [Hawthorne, 2002: 110]. By contrast with Hawthorne, however, I note that controversies come in degrees, and I further suppose that, the more incoherent an option is, the less controversial it should be to dismiss it as an impossibility. This leads me to judge that, since magic is *highly* incoherent, to that extent we should dismiss it as an impossibility. As such, non-physical blockers should be dismissed as impossible.

<sup>5</sup> Hawthorne, 2002: 105.



Since non-physical blockers are impossible, Hawthorne's blocker arguments have no—or exceedingly minimal—force unless, by coincidence, he is right that any tenable formulation of physicalism should incorporate PD. He is right that physicalism should imply a physical dependence principle of some sort. Yet, since non-physical blockers are impossible, PD is presumably overkill. In its place, I argue that the following physical dependence principle, which doesn't imply PD, is strong enough:

PD\* Physicalism is true only if every (instantiated) micro property is microphysical.

PD\* captures the physical dependence intuition that materialistic worlds cannot have truths that depend on the instantiation of micro-*non-physical* properties. I doubt a definition of physicalism should have to imply a stronger physical dependence principle than this. If I am right that any tenable formulation of physicalism should imply PD\*, then  $D_{jac}$  and  $D_{cha}$ , if they are tenable, should imply PD\*. I will now show they imply PD\*.

$D_{jac}$  implies PD\*. For PD\* is equivalent to the thesis that physicalism is false of any world that has micro-nonphysical properties. Yet  $D_{jac}$  implies this. After all, for any world that has micro-nonphysical properties, there will be a *minimal* microphysical duplicate world that lacks micro-nonphysical properties and thus isn't a duplicate simpliciter.

The only remaining question is whether PD\* is strong enough to capture *all* of the right physical dependence intuitions about physicalism. Yet, given the plausibility that non-physical blockers are impossible, it is difficult to see what other physical dependence intuitions should be captured. Thus PD\* seems sufficient; thus  $D_{jac}$  seems to imply all the right physical dependence principles. Thus  $D_{jac}$  seems safe from Hawthorne's blocker critique.

What about  $D_{cha}$  and  ${}_rD_{cha}$ ? Do they imply all the right physical dependence principles? Like  $D_{jac}$ ,  $D_{cha}$  implies PD\*. For PD\* is equivalent to the thesis that physicalism is false of any world that has micro-nonphysical properties. Yet  $D_{cha}$  implies this. After all, for any world  $w$  that has micro-nonphysical properties, there will be a microphysically indiscernible world that does not have all of  $w$ 's positive truths: at any rate, I just argued no *minimal* microphysical duplicate will have non-physical properties.

Meanwhile,  ${}_rD_{cha}$  does not imply PD\*, as  ${}_rD_{cha}$  restricts its claim to just the mental facts instead of all the facts. Yet  ${}_rD_{cha}$  does imply the following restricted version of PD\*:

${}_rPD^*$  Physicalism is true of the mental only if every basic property (upon which some mental property depends (wholly or partly)) is microphysical.

${}_rD_{cha}$  does not seem to need a physical dependence principle stronger than  ${}_rPD^*$ , and yet  ${}_rD_{cha}$  implies  ${}_rPD^*$ . For  ${}_rPD^*$  says that physicalism is false of the mental in any world where there are micro-nonphysical proper-



ties upon which some mental properties depend.  ${}_rD_{cha}$  implies this. After all, for any world  $w$  that has micro-nonphysical properties upon which some mental properties depend, there will be a microphysically indiscernible world that does not have all of  $w$ 's positive mental truths. For there will at least be a minimal microphysical duplicate of  $w$  that will be microphysically indiscernible from  $w$  but won't have  $M$ , since, if it did, it wouldn't be a *minimal* microphysical duplicate. Thus a minimal microphysical duplicate of  $w$  wouldn't have  $w$ 's positive mental truths that depend on its micro-nonphysical properties, since a minimal microphysical duplicate wouldn't have any micro-nonphysical properties.

Altogether,  ${}_rPD^*$  gives a necessary condition of physicalism of the mental; but any counterexample to physicalism of the mental on  ${}_rPD^*$  would be a counterexample to physicalism of the mental on  ${}_rD_{cha}$ ; thus  ${}_rPD^*$  is a necessary condition of  ${}_rD_{cha}$ ; i.e.,  ${}_rD_{cha}$  implies  ${}_rPD^*$ .

Alas, there is one further consideration to strengthen my critique of Hawthorne's blocker problem for  ${}_rD_{cha}$ . Namely, even if one is interested in restricted physicalisms, I argue that this doesn't require giving *restricted formulations* of physicalism. On the one hand, perhaps some conglomerations of properties can be instantiated in worlds where they are the only instantiated properties in the world. That is, perhaps there can be a world consisting of precisely that conglomeration of instantiated properties and nothing else. In such a case, since  $D_{cha}$  defines physicalism of a *world*, it thereby defines physicalism of that restricted class of instantiated properties, since it defines physicalism of a world consisting of just that conglomeration of instantiated properties. On the other hand, even if some conglomerations of properties cannot be instantiated in worlds that have no other instantiated properties, it is not clear that such conglomerations deserve a restricted formulation of physicalism that gives conditions for whether physicalism is true of them alone. For, in so far as they cannot exist in a world all by themselves, they are necessarily connected in some way to other properties. Thus, if we are going to ask whether physicalism is true of them, presumably we should instead ask whether physicalism is true of a minimal conglomeration that they are connected with. But, in so far as a minimal conglomeration is defined as needing to be able to be instantiated all by itself in a world of its own, then once again we don't need a restricted formulation of physicalism; all we need is a formulation of physicalism of a world.

Altogether,  $D_{cha}$  implies a physical dependence principle,  $PD^*$ , that seems strong enough to capture appropriate physical dependence intuitions with respect to physicalism's truth of a world. Meanwhile, restricted formulations of physicalism are unnecessary since any minimal conglomeration of properties could exist in a world of its own anyway and therefore be covered by a definition of physicalism of a world. Furthermore, even if restricted formulations of physicalism were necessary,  ${}_rD_{cha}$  implies  ${}_rPD^*$ , which seems to capture the right physical dependence intuitions with respect to physicalism's truth of the mental.



## Conclusion

Hawthorne's blocker arguments serve as problems for the two most popular ways of formulating physicalism rigorously: Frank Jackson's formulation,  $D_{jac}$ , and David Chalmers' formulation,  $D_{cha}$ . Yet his blocker arguments are misguided since non-physical blockers are impossible, and his blocker argument against  $D_{cha}$  is unsound since it falsely presupposes that restricted materialisms require restricted formulations of materialism. Finally, he is wrong to think  $D_{jac}$  and  $D_{cha}$  don't capture the right physical dependence principles. For  $D_{jac}$  and  $D_{cha}$  each imply the physical dependence principle,  $PD^*$ , which, given the impossibility of non-physical blockers, is a sensible replacement to  $PD$ ; likewise, even if restricted formulations of physicalism are required,  ${}_rD_{cha}$  implies the restricted physical dependence principle,  ${}_rPD^*$ , which, given the impossibility of non-physical blockers, is a sensible replacement to  ${}_rPD$ . In these ways, Hawthorne's blocker arguments have been blocked.

## Acknowledgements

I would like to thank René Jagnow and Igor Gasparov for their very insightful comments and advice that helped me improve this paper.

## References

- Chalmers, 1996 – *Chalmers D.* The Conscious Mind. Oxford : Oxford University Press, 1996.
- Chalmers, Jackson, 2001 – *Chalmers D., Jackson F.* Conceptual Analysis and Reductive Explanation // *Philosophical Review*. 2001. Vol. 110, № 3. P. 315–361.
- Hawthorne, 2002 – *Hawthorne J.* Blocking Definitions of Materialism // *Philosophical Studies*. 2002. Vol. 110, № 2 (March). P. 103–113.
- Jackson, 1982 – *Jackson F.* Epiphenomenal Qualia // *The Philosophical Quarterly*. 1982. Vol. 32, № 127 (April). P. 127–136.
- Jackson, 1986 – *Jackson F.* What Mary Didn't Know // *The Journal of Philosophy*. 1986. Vol. 83, № 5 (May). P. 291–295.
- Jackson, 1994 – *Jackson F.* Armchair Metaphysics // *Philosophy in Mind*; eds. M. Michael and J. O'Leary-Hawthorne. Dordrecht, The Netherlands : Kluwer Academic Publishers, 1994. P. 23–42.
- Jackson, 1998 – *Jackson F.* From Metaphysics to Ethics: A Defence of Conceptual Analysis. Oxford : Oxford University Press, 1998.
- Stoljar, 2009 – *Stoljar D.* Physicalism // *Stanford Encyclopedia of Philosophy*. 2009.



## П ОЧЕМУ РАЗРЫВ В ОБЪЯСНЕНИИ НЕ ЯВЛЯЕТСЯ РЕШАЮЩИМ АРГУМЕНТОМ ПРОТИВ НАТУРАЛИЗМА В ФИЛОСОФИИ СОЗНАНИЯ<sup>1</sup>

**Секацкая Мария Александровна** – ст. преподаватель кафедры философии науки и техники Санкт-Петербургского государственного университета.  
E-mail: maria.sekatskaya@gmail.com



Явления сознательной жизни субъективны, мы наблюдаем их из перспективы первого лица. Наука стремится к объективности, ее теории рассуждают о явлениях, доступных всеобщему наблюдению, т.е. как бы наблюдаемых из перспективы третьего лица. В философии сознания существует несколько аргументов и соображений, призванных показать, что объективный научный подход не способен адекватно объяснить сознание. В статье я кратко описываю эти соображения и подробно рассматриваю знаменитый аргумент Дж. Левина о существовании «разрыва в объяснении» между явлениями сознательной жизни, как мы наблюдаем их из перспективы первого лица, и научными теориями, призванными объяснить эти явления. Я привожу аргументы в пользу того, что существование «разрыва в объяснении» не может служить доказательством против физикализма или функционализма, поскольку этот разрыв является формулировкой той же самой интуиции, на которой основываются мысленные эксперименты антифизикалистов и антифункционалистов. Далее я показываю, что «разрыв в объяснении» характерен не только для научных теорий сознания, но для любых научных теорий, и делаю вывод, что он не является аргументом в пользу антинатурализма в философии сознания.

**Ключевые слова:** разрыв в объяснении, квалиа, субъективный опыт, физикализм, функционализм, философские зомби.

## WHY THE EXPLANATORY GAP IS NOT A DECISIVE ARGUMENT AGAINST NATURALISM IN PHILOSOPHY OF MIND

**Maria Sekatskaya** – PhD. Senior lecturer, chair of philosophy of science, Saint-Petersburg State University.

Conscious phenomena are subjective and are experienced from the first person perspective. Science aims for objectivity. Scientific theories deal with publicly observable phenomena, which are therefore said to be given from the “third person perspective”. In philosophy of mind there have been presented several arguments to the effect that objective scientific approach can’t explain consciousness. In this article I focus on the famous argument of “explanatory gap” by Joseph Levine. According to Levine, there is an unremovable gap in any scientific explanation of the phenomena of conscious experience, because no such theory can intelligibly correlate objective scientific concepts with subjective phenomenal states. I argue for two claims.

First, I want to show that “explanatory gap” can not be used as an independent argument against physicalism or functionalism such that it would give anti-physicalist or anti-functionalist thought experiments additional conceptual force. It can’t be used for this aim not because it is wrong, but because it is not an argument, strictly speaking, but rather a formal expression of basic anti-physicalist and anti-functionalist intuitions.

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ, проект № 13-06-00902.



Second, I want to demonstrate that “explanatory gap” is not specific for explaining consciousness. This kind of gap also exists in other theoretical explanations of concrete facts and is not considered problematic in those other cases. In order to show this I argue that physical theories of space can not explain why I am “here” and physical theories of time can not explain why I am “now”, but we do not think that they leave out something important for our understanding of space and time. My conclusion is that the existence of “explanatory gap” is not a decisive argument in favor of anti-naturalism in philosophy of mind.

**Key words:** explanatory gap, qualia, consciousness, experience, physicalism, functionalism, philosophical zombies.

Большинство участников дебатов в философии сознания согласно с тем, что слова, указывающие на события в сознании, и слова, указывающие на события в так называемой «объективной действительности», или в «физической реальности», частью которой является головной мозг человека, переживающего эти состояния сознания, имеют разное значение<sup>2</sup>. Вопрос в том, как интерпретировать этот факт. Я хочу показать, что несмотря на то, что слова, обозначающие феноменальные состояния сознания, и слова, обозначающие процессы в «физической реальности», имеют разное значение, и несмотря на то, что мы по-разному убеждаемся в их истинности или ложности, возникающий из этого положения дел «разрыв в объяснении» не является аргументом в пользу антинатурализма в философии сознания.

Как отмечает Фреге, из того, что два слова имеют разные значения, не следует, что они указывают на разные предметы. Так, слова «Утренняя Звезда» и «Вечерняя Звезда» означают разное, но указывают на один и тот же предмет – планету Венера [Фреге, 2000]. После того как Фреге поставил вопрос таким образом, и началась, на мой взгляд, современная эпоха в дебатах об онтологическом статусе сознания. Из того, что из описания с использованием слов «я ощущаю красное» невозможно аналитически вывести описание с использованием слов «в моем мозгу определенная популяция нейронов зоны зрительной коры находится в состоянии возбуждения», еще не следует, что эти два способа описания указывают на разное состояние дел. Возможно, состояние дел одно и то же, но мы по-разному удостоверяемся в нем – непосредственно, или «изнутри», т.е. испытывая это состояние, и опосредованно, или «снаружи», т.е. наблюдая того, кто испытывает это состояние [Плейс, 1956].

Но если два способа описания указывают на один и тот же объект, или состояние дел, и если мы можем каким-то образом удостовериться в том, что они оба истинны, должен быть способ установить между ними взаимно-однозначное соответствие. Более того, если один способ описания применим ко всем объектам природы, а второй – только к объектам,

<sup>2</sup> См., напр., Фреге Г. Мысль. Логическое исследование [Фреге, 2000]; Kripke S. Naming and Necessity [Крипке, 1980]; Place U.T. Is Consciousness a Brain Process? [Плейс, 1956]; Levine J. Materialism and Qualia: the Explanatory Gap [Левин, 1983].



## ПОЧЕМУ РАЗРЫВ В ОБЪЯСНЕНИИ НЕ ЯВЛЯЕТСЯ...

обладающим сознанием, то такое взаимно-однозначное соотношение будет представлять собой редукцию ментального к физическому, потому что теория с меньшей объяснительной силой станет частным случаем теории с большей объяснительной силой [Льюис, 1994].

В связи с этим, как мне кажется, противники редукционизма уделяют такое внимание различным мысленным экспериментам, призванным показать, что мы не можем установить взаимно-однозначное соответствие между событиями ментального мира и событиями физического мира. Признавая объяснительную силу редукционистского анализа для всех практических целей, антиредукционисты могут настаивать на том, что он оставляет что-то за скобками. Предположим, что философу-редукционисту удалось описать все квалиа в функциональных терминах или установить взаимно-однозначное соответствие между состояниями сознания и состояниями мозга. Но антиредукционисты по-прежнему будут настаивать, что такое совпадение случайно. Мы ведь можем представить себе, что существует система, которая выполняет те же функции, но при этом лишена квалиа [Блок, 1978]. Мы также можем представить себе, что система имеет ту же физическую структуру, но при этом лишена квалиа [Чалмерс, 1995, 1996]. Такие аргументы кажутся интуитивно очевидными<sup>3</sup>.

И здесь дебаты заходят в тупик. Если мы можем представить себе, что любое, даже самое полное функциональное и физическое описание системы не дает оснований для вывода о том, что у этой системы имеются (или не имеются) определенные состояния сознания, то редукция ментального описания к физическому тем самым оказывается незавершенной. Редукционисты могут настаивать, что каждый конкретный мысленный эксперимент, приводимый сторонниками антиредукционизма, некорректен или не доказывает того, что должен доказать. Это не является решающим контраргументом с точки зрения антиредукционистов, потому что непоколебленной остается главная интуиция: есть нечто такое, что неподвластно описанию в физических и функциональных терминах. И если так, то описание в этих терминах является неполным. Из этого тезиса можно выводить более или менее фундаментальные следствия. Дэвид Чалмерс полагает, что из мыслимости зомби следует, что сознание есть физически фундаментальная сущность наряду с временем и пространством [Чалмерс, 1995]. Джозеф Левин предпочитает придерживаться более слабого тезиса, который состоит в том, что в любой попытке редукции ментального к физическому присутствует «разрыв в объясне-

<sup>3</sup> Дэниел Деннет, посвятивший много усилий объяснению квалиа в функциональных терминах, называет эту интуицию «предчувствием зомби» – *zombic hunch*, и отмечает, что побороть ее рациональными аргументами очень сложно, если не невозможно. См.: *Dennett D. Sweet Dreams: Philosophical Obstacles to a Science of Consciousness* [Деннет, 2005].



нии» (explanatory gap). В статье «Материализм и квалиа: разрыв в объяснении» он ссылается на мысленные эксперименты Блока и Патнема, призванные продемонстрировать, что описания в качественных терминах не выводимы из физических и функциональных описаний [Левин, 1983]. Левин аргументирует следующим образом: отождествление физического описания состояний мозга и нервной системы с качественным описанием состояний сознания будет иметь вид:

(1) боль – это возбуждение С-волокон<sup>4</sup>;

отождествление одного физического описания с другим будет иметь вид:

(2) тепло – это движение молекул;

отождествление функционального описания состояний мозга (или организма) с качественным описанием состояний сознания будет иметь вид:

(3) испытывать боль – означает быть в состоянии Ф.

Есть ли что-то, различающее (1), (2) и (3) с формальной точки зрения? Что-то такое, что обосновывает имеющееся у нас «предчувствие зомби»?

Левин говорит следующее: «В нашем понятии боли содержится нечто большее, чем представление о каузальной роли, которую она играет. Есть также ее качественный характер, то, каково это – ощущать боль. То, что остается необъясненным после открытия возбуждения С-волокон, – это *почему боль ощущается так, как она ощущается!* Ведь кажется, что в возбуждении С-волокон нет ничего такого, что делало бы его естественным образом более «подходящим» (fit) для феноменальных качеств боли, чем для любого другого набора феноменальных качеств. В отличие от отождествления функциональной роли боли с возбуждением С-волокон (или с каким-то свойством такого возбуждения), отождествление качественной стороны боли с таким возбуждением оставляет природу их связи совершенно загадочной. Можно сказать, что в результате то, как именно ощущается боль, оказывается примитивным, неанализируемым фактом (brute fact)» [Левин, 1983: 357].

Но почему этот факт должен поддаваться анализу? В конце концов, не всякий факт может быть далее проанализирован и не на всякий вопрос о тождестве можно дать информативный ответ. Например, на вопрос о том, почему стол, на котором стоит мой компьютер, тождествен самому себе, а не чему-то другому, информативного ответа быть не может.

<sup>4</sup> Это отождествление не претендует на полноту – в полном отождествлении помимо возбуждения С-волокон будут фигурировать возбуждение нейронов специфических отделов головного мозга и т.д. Левин указывает, что такое краткое отождествление осуществляется для удобства и не влияет на аргументацию.





## ПОЧЕМУ РАЗРЫВ В ОБЪЯСНЕНИИ НЕ ЯВЛЯЕТСЯ...

В статье «Материализм и квалиа: разрыв в объяснении» Левин приводит несколько соображений о том, почему тождества типа (1) и (3) представляются требующими дальнейшего анализа. Левин отмечает, что проблема редукции ментального не в том, что ее нельзя провести по правилам, предлагаемым номологическо-дедуктивной моделью объяснения. Левин согласен, что редукция может соответствовать этим правилам; тем не менее, она не сделает такое объяснение понятным: «Нам требуется некое представление о том, что означает, что какой-то феномен был сделан *понятным* (intelligible), вместе с правилами, определяющими, когда требование большей понятности становится неуместным. Например, я полагаю, что законы гравитации объясняют в том самом смысле, который мы ищем здесь, феномен падения тел. Ничто в этом объяснении не кажется оставленным за скобками. Тем не менее говорится, что значение  $G$  гравитационной константы не выводится ни из каких базовых законов. Это данный, примитивный, неанализируемый факт о вселенной. Остается ли у нас ощущение, что что-то, нуждающееся в объяснении, не было объяснено? Или мы ожидаем, что некоторые факты природы должны казаться случайными в этом смысле? Я склонен согласиться с последним утверждением относительно  $G$ . Но тогда можно спросить, почему связь между тем, каково это, пребывать в определенном функциональном (или физическом) состоянии и самим этим состоянием требует объяснения для того, чтобы быть понятной?» [Левин, 1983: 358].

Можно сказать, что этот вопрос является вопросом об обоснованности «предчувствия зомби», т.е. обоснованности нашей интуиции о том, что редукция ментального к нементальному всегда будет неполной. Какие же аргументы в пользу его обоснованности выдвигает Левин?

1. Феномен сознания возникает на макроскопическом уровне — только высокоорганизованные физические системы демонстрируют наличие ментальности. Это вполне соответствует тому, что функционалисты говорят о сознании. Однако, пишет Левин, «кажется странным, что примитивные факты того рода, которые представлены в утверждениях типа (1) и (3), возникают на этом уровне организации. Материализм, как я его понимаю, требует по крайней мере минимального объяснительного редукционизма такого рода: для каждого феномена, не описываемого в терминах фундаментальных физических величин (чем бы они не оказались), существует такой механизм, описываемый в терминах фундаментальных физических величин, что проявления первого механизма объясняются в терминах проявлений второго (occurrences of the former are intelligible in terms of occurrences of the latter). И хотя этот минимальный редукционизм ничего не говорит о сводимости теорий, например психологии к физике, он подразумевает, что примитивные факты, подобные факту о числовом значе-



нии G, не возникают в сфере теорий, подобных психологии» [Левин, 1983: 358–359].

Таким образом, Левин говорит, что для полного объяснения ментального нам недостаточно установления взаимно-однозначного соответствия между двумя типами описания – в терминах ментального и в терминах физического. Нам нужно еще некоторое объяснение факта тождества между этими описаниями. Однако не совсем понятно, что имеется в виду под примитивными фактами, проявляющимися на разных уровнях. Если странно, что факты ментального проявляются только на уровне высокоорганизованных физических систем, будет ли менее странно, если они будут проявляться на более фундаментальном уровне? Например, будет ли менее странно, если выяснится, что за проявление ментальных качеств ответственны какие-то базовые кирпичики реальности, скажем, М-частицы? Или М-поля? Или М-струны? Это утверждение представляется как минимум неочевидным, не говоря уже о том, что оно приведет ко всем тем проблемам, что знакомы нам относительно других отождествлений. Например, предположим, что мы обнаружили, что во всех случаях проявления ментальных событий на уровне физического описания системы, к которой относятся эти ментальные события, мы можем обнаружить некие фундаментальные М-частицы (или М-поля). Достаточно ли это для того, чтобы утверждать тождество между М-частицами и сознанием, или мы можем представить себе зомби-систему, аналогичную этой системе и обладающую такими же М-частицами, но лишенную при этом каких бы то ни было ментальных качеств? Если можем, то связь между М-частицами и ментальными качествами ничуть не менее необъяснима, чем связь между функциональными состояниями и ментальными качествами. Я полагаю, что этих соображений достаточно, чтобы показать, что тот факт, что явления сознания проявляются на высоком уровне организации физических систем, не кажется в большей степени необъяснимым, чем если бы они проявлялись на самом фундаментальном уровне. И в том, и в другом случае мы можем испытывать «предчувствие зомби».

2. Второй аргумент в пользу антиредукционистской интуиции, который выдвигает Левин, заключается в следующем. Разрыв в объяснении – причина того, что тождества типа (1) и (3) представляются нам случайными в отличие от тождеств типа (2). Даже если какие-то утверждения о психофизическом тождестве являются истинными, мы никогда не узнаем, какие именно. Предположим, боль тождественна какому-то физическому состоянию, а именно: возбуждению С-волокон, потому что именно с возбуждением С-волокон коррелирует боль как минимум в случае, когда ее испытывает человек. Тогда, если мы обнаружим инопланетян, высказывающих все функциональные признаки боли, но лишенных С-волокон (функцию С-волокон выполняют у них D-клапаны), должны ли мы заключить, что они не



## ПОЧЕМУ РАЗРЫВ В ОБЪЯСНЕНИИ НЕ ЯВЛЯЕТСЯ...

испытывают квалитативного чувства боли, или мы должны предположить, что боль тождественна дизъюнкции двух разных физических состояний – возбуждения С-волокон или открытия D-клапанов? Левин отмечает, что этот аргумент, используемый в дебатах между физикалистами (сторонниками теории тождества) и функционалистами, он приводит не для того, чтобы показать, что правы функционалисты, а для того чтобы продемонстрировать, что любое отождествление квалитативных состояний с физическими или функциональными состояниями оставляет разрыв в объяснении незаполненным: «Какие возможные свидетельства помогут нам выбрать между двумя гипотезами: что боль есть возбуждение С-волокон или что боль есть либо возбуждение С-волокон, либо открытие D-клапанов?

Ведь если бы можно было различить какую-то внутреннюю связь между тем, что чьи-то С-волокна находятся в состоянии возбуждения (или что кто-то пребывает в функциональном состоянии Ф), и тем, каково это – ощущать боль, т.е. если бы ощущения можно было понять в физических или функциональных терминах, то мы могли бы вывести меру сходства [наших ощущений и ощущений инопланетян. – М.С.] из природы этого объяснения. Те качества возбуждения С-волокон (или пребывания в состоянии Ф), которые объясняли бы ощущение боли, определяли бы, на какие особенности физического (или функционального) описания системы мы должны обращать внимание. Но пока не заполнен разрыв в объяснении, мы не можем обнаружить факты, определяющие наличие и квалитативный характер ощущения боли у существ, физически (или функционально) отличных от нас. Это в свою очередь влечет за собой то, что истинность или ложность утверждений типа (1) хотя и является метафизическим фактом, эпистемически недоступна для нас» [Левин, 1983: 360].

Я бы хотела отметить, что второй аргумент Левина в пользу антиредукционистской интуиции подразумевает, что взаимно-однозначное тождество между квалитативными состояниями и физическими (функциональными) состояниями установить невозможно. В самом деле, если установить его возможно, то ложен вывод Левина о том, что истинность или ложность утверждений типа (1) эпистемически недоступна для нас (а если ложен вывод, то опровергнут и сам аргумент). Следовательно, он не может быть использован для подтверждения истинности антиредукционистской интуиции без *petitio principia*.

Таким образом, как мне кажется, второй аргумент Левина является не доказательством истинности (или теоретической оправданности) антиредукционистской интуиции, а ее формальным изложением. Если такие философы, как Дэвид Чалмерс и Фрэнк Джексон, изобретают мысленные эксперименты, призванные показать, что из физического (или функционального) описания системы нельзя вывести заключения о наличии у нее квалитативных состояний, Джозеф Левин



обобщает этот способ аргументации в своем рассуждении о том, что между физическим (функциональным) описанием и качественным описанием сознания всегда останется разрыв в объяснении.

Я полагаю, однако, что этот разрыв в объяснении не является специфической проблемой для философии сознания, а потому он не позволяет утверждать, что сознание в большей степени необъяснимо, чем другие явления природы. Как мне кажется, это утверждение можно продемонстрировать следующими соображениями.

Описание сознания в терминах физической структуры или функциональной роли представляет собой научную теорию, подобную другим научным теориям<sup>5</sup>. Существует множество конкурирующих научных теорий сознания, пользующихся этим способом описания, – теория А. Дамазио, теория Дж. Эдельмана, теория Р. Ллинаса, теория С. Дехейне и Л. Накаша и т.д.<sup>6</sup> Несмотря на то что окончательная Единая Теория Сознания на данный момент не сформулирована, ситуация в этой области естественных наук принципиально не отличается от ситуации в других областях естественных наук, например в физике. В самом деле несмотря на то, что теория квантовой гравитации до сих пор не разработана, и мы не знаем, как именно объединить общую теорию относительности с квантовой механикой, мало кто рассматривает это как повод занять антинатуралистическую позицию относительно пространства и времени. Таким образом, тот факт, что единая теория сознания не сформулирована, не может сам по себе быть аргументом в пользу дуализма. Однако Джозеф Левин говорит не об этом. Вопрос, который он ставит, заключается не в том, как дать единую и полную естественно-научную теорию сознания, наилучшим образом объясняющую все известные экспериментальные данные из области нейрофизиологии, клинической психологии, когнитивных наук и т.д., а в том, как объяснить, почему данное физическое (или функциональное) теоретическое описание соответствует данному состоянию сознания.

Но как объясняется «данное» состояние сознания? Как мы узнаем, о каком именно квале идет речь? Путем остенсивного определения. Если мы посмотрим на сотни статей, посвященных проблеме качественных состояний сознания, то определение этих состояний дается через указание. Сознание описывается при помощи отсылки к конкретному индивидуальному опыту – «ощущение, которое я ис-

<sup>5</sup> Д. Деннет называет методологию, применимую для соотношения объективных и субъективных описаний, применимую в любой научной теории сознания, «гетерофеноменологией» и показывает, как она может работать: *Dennett D.C. Consciousness Explained* [Деннет, 1991].

<sup>6</sup> См.: *Damasio A. The Feeling of What Happens* [Дамазио, 1999]; *Edelman, G.M. The Remembered Present: A Biological Theory of Consciousness* [Эдельман, 1990]; *Llinas R. I of the Vortex: From Neurons to Self* [Ллинас, 2001]; *Dehaene S., Naccache L. Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework* [Дехейн, Накаш, 2001].



## ПОЧЕМУ РАЗРЫВ В ОБЪЯСНЕНИИ НЕ ЯВЛЯЕТСЯ...

пытываю, вдыхая аромат свежемолотого кофе», или «слушая Девятую симфонию Бетховена», или «пробуя шоколад», или «прикасаясь к оголенному проводу под напряжением». Иначе говоря, такие определения на самом деле подобны указаниям: вот это! И эпистемологический разрыв состоит в том, чтобы объяснить, как связаны «вот это» ощущение и его теоретическое описание.

Однако та же самая проблема стоит перед другими способами научного объяснения, которые в силу каких-то причин вызвали значительно меньший интерес у авторов, пишущих про философию сознания. Я бы хотела исправить эту ситуацию и провести соответствующие параллели.

Например, мне кажется, что теория пространства-времени Эйнштейна (и даже теория струн!) не объясняет, почему я нахожусь именно здесь и почему данный момент – это именно сейчас. Конечно, они объясняют, что такое пространство, и что такое траектория в пространстве-времени, и что я, если рассматривать меня как четырехмерный объект, в каждый последующий момент времени занимаю определенную точку, задаваемую при помощи трех пространственных и одной временной координаты (в случае теории струн потребуется больше координат). Однако они не объясняют, чем «здесь» отличается от «там», а «сейчас» отличается от «тогда». Иными словами, они не объясняют, чем одна точка этого четырехмерного объекта, которая имеет место здесь и сейчас, отличается от других точек этого объекта, находящихся там и тогда.

Подобно тому, как научная теория сознания (например, теория Р. Линаса или Дж. Эдельмана) объясняет процессы, происходящие в мозге, в их общем виде, например «возбуждение световых рецепторов на сетчатке глаза ведет к передаче возбуждения по оптическому нерву, которое вызывает возбуждение определенных популяций нейронов в зонах зрительной коры V1–V4, которое ведет к возбуждению определенной популяции нейронов в области мозга, контролирующей произвольные движения, что ведет к возбуждению интернейронов, вызывающих движение руки» (несколько важных стадий описания опущено для краткости), так и физическая теория описывает процессы в общем виде, например: сила равна произведению массы на ускорение ( $F = ma$ ), энергия равна произведению массы на скорость света в квадрате ( $E = mc^2$ ) и т.д. Ни из одной, самой полной физической теории невозможно сделать вывод, что описываемый процесс происходит «здесь и сейчас», так как физические теории в равной степени применимы ко всем областям пространства и времени (при условии, что в этих областях соблюдаются те же самые физические законы).

Иными словами, разрыв в объяснении, о котором говорит Левин, это не особая проблема теорий сознания, а разрыв между теорией и фактами. Теория не объясняет факт как таковой – она связывает



факты между собой. Самая лучшая теория не поможет нам ничего объяснить, если мы не зададим какое-то положение дел, относительно которого мы хотим предсказать будущее или вычислить прошлое. Теория имеет объяснительную силу постольку, поскольку она способна подводить фактические наблюдения под общие закономерности [Гемпель, 1965]. Подобно тому, как теория пространства-времени не должна объяснять, чем «здесь» отличается от «там», а должна давать общие законы перемещения тел в пространстве, так и теория сознания не должна объяснять, почему ощущение зеленого «вот такое» (зеленое квале), а ощущение красного – «вот такое» (красное квале); она должна лишь объяснить, как эти ощущения связаны друг с другом и прочими фактами действительности.

Следовательно, даже если тождества типа (1) и (3) кажутся нам «случайными» и мы всегда можем представить себе, что точно такой же физический субстрат (или функционально аналогичная система) лишены сознания, «вот этого квале», из этого не следует, что теория сознания неполна. Иначе неполна и наша теория пространства, как, я надеюсь, покажет мой аргумент о «разрыве в объяснении» в физических теориях пространства. Аргумент таков.

Физика объясняет, по каким законам перемещаются тела в пространстве и во времени. Я являюсь физическим телом. Мои перемещения могут быть описаны по законам физики. Мое положение в пространстве может быть определено по отношению к другим телам. При этом из перспективы первого лица я знаю о себе, что нахожусь здесь. Представим себе, что в другом возможном мире какая-нибудь удаленная галактика расположена иначе (в силу квантовой флуктуации в момент Большого взрыва). В этом возможном мире все точно так же, за исключением этой удаленной галактики. Там даже есть я\*, точнее, мой двойник из возможного мира, имеющий точно такие же объективные свойства и субъективные ощущения, как я в актуальном мире. С точки зрения физической теории пространства-времени, мое\* положение в пространстве в этом возможном мире будет отличаться от моего положения в пространстве в актуальном мире. Но из перспективы первого лица я\* из возможного мира все равно буду считать, что нахожусь «здесь». При этом мои ощущения будут тождественны и мое убеждение о том, что я нахожусь «здесь», будет истинным как для актуального мира, так и для этого возможного мира, где я\* также считаю, что нахожусь «здесь». При этом как я из актуального мира, так и я\* из возможного мира будем правы, когда будем считать, что утверждение «я нахожусь здесь» истинно. Однако с точки зрения физики мое положение и мое\* положение в пространстве будут разными. Следовательно, тот факт, что я нахожусь здесь, необъясним исключительно при помощи понятийного аппарата физики. Должны ли мы сделать вывод, что физическая теория пространст-



## ПОЧЕМУ РАЗРЫВ В ОБЪЯСНЕНИИ НЕ ЯВЛЯЕТСЯ...

ва-времени принципиально неполна, пока она не объяснит, что значит «здесь»?

Я полагаю, что если этот аргумент не убеждает нас, что разрыв в объяснении является достаточным основанием для антинатурализма и в сфере исследования природы, то и аргумент Левина не может рассматриваться как достаточное основание для антинатурализма в сфере исследования сознания. То, что Левин называет «разрывом в объяснении», – это разрыв между фактами сознания и теорией сознания, объясняющей эти факты, поэтому заполнить его невозможно, ведь подобное заполнение будет либо еще одним необъяснимым фактом, либо теорией, подводящей этот факт под общую закономерность, а любое такое подведение оставит «разрыв в объяснении» незаполненным.

## Библиографический список

- Блок, 1978 – *Block N.* Troubles with Functionalism // Minnesota Studies in The Philosophy of Science. 1978. P. 261–325.
- Гемпель, 1965 – *Hempel C.* Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. N.Y. : Free Press, 1965.
- Дамазио, 1999 – *Damasio A.* The Feeling of What Happens. N.Y. : Harcourt Brace, 1999.
- Деннет, 1991 – *Dennett D.C.* Consciousness Explained. Boston: Little, Brown and Company, 1991.
- Деннет, 2005 – *Dennett D.C.* Sweet Dreams: Philosophical Obstacles to a Science of Consciousness. MIT Press, 2005.
- Дехейн, Накаш, 2001 – *Dehaene S., Naccache L.* Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework // Cognition. 2001. Vol. 79 (1–2). P. 1–37.
- Крипке, 1980 – *Kripke S.* Naming and Necessity. Cambridge (Mass.), 1980.
- Левин, 1983 – *Levine J.* Materialism and Qualia: the Explanatory Gap // Pacific Philosophical Quarterly. 1983. Vol. 64. P. 354–361.
- Ллинас, 2001 – *Llinas R.* I of the Vortex: from Neurons to Self. Cambridge, MA: MIT Press, 2001.
- Льюис, 1994 – *Lewis D.* Reduction of Mind / S. Guttenplan (ed.). A Companion to Philosophy of Mind. Oxford : Blackwell Publishers, 1994. P. 412–431.
- Плейс, 1956 – *Place U.T.* Is Consciousness a Brain Process? // British Journal of Psychology. 1956. Vol. 47. P. 44–50.
- Фреге, 2000 – *Фреге Г.* Мысль. Логическое исследование // Г. Фреге. Логика и логическая семантика. М. : Аспект-Пресс, 2000; *Kripke S.* Naming and Necessity. Cambridge (Mass.), 1980.
- Чалмерс, 1995 – *Chalmers D.* Facing Up to the Problem of Consciousness // Journal of Consciousness Studies 1995. Vol. 2 (3). Электронная публикация <http://www.imprint.co.uk/chalmers.html>.
- Чалмерс, 1996 – *Chalmers D.* The Conscious Mind, Oxford: Oxford University Press, 1996.
- Эдельман, 1990 – *Edelman G.M.* The Remembered Present: A Biological Theory of Consciousness. N.Y. : Basic Books, 1990.



## П ОЧЕМУ НАС ДОЛЖЕН ЗАБОТИТЬ РАЗРЫВ В ОБЪЯСНЕНИИ СОЗНАНИЯ, А НЕ ТРУДНАЯ ПРОБЛЕМА СОЗНАНИЯ?

**Светлана  
Фарвазовна Нагума-  
нова** – докт. филос.  
наук, доцент кафедры  
истории, философии,  
социологии и политоло-  
гии Казанского госу-  
дарственного медицин-  
ского университета.  
E-mail:  
nagouman@mail.ru

В философии сознания сегодня принято считать, что трудная проблема сознания и проблема разрыва в объяснении сознания – это одна и та же проблема. Понятие разрыв в объяснении ввел Д. Левин (1983). Д. Чалмерс (1995) развил идею Левина в трудную проблему сознания. В этой статье я показываю, что здесь мы имеем дело с двумя разными проблемами, и обосновываю, почему нас должен заботить разрыв в объяснении сознания, а не трудная проблема. И Левин, и Чалмерс принимают функциональную модель объяснения. Они считают, что функциональная редукция, которая так хорошо работает для почти всех явлений, не работает в объяснении сознания. Ни одна когнитивная функция не может схватить феноменальное сознание. Тем не менее функциональная модель допускает компромисс: возможность замены нашего дотеоретического понятия сознания его теоретическим понятием-преемником. Наиболее подходящим понятием-заменой, на мой взгляд, является понятие феноменальной репрезентации.

**Ключевые слова:** сознание, Трудная проблема, разрыв в объяснении, функциональная редукция, замена, понятие-преемник, феноменальная репрезентация.

## W HY SHOULD WE CARE ABOUT THE EXPLANATORY GAP, AND NOT THE HARD PROBLEM?

**Svetlana Nagumanova** –  
Kazan State Medical  
University



It is a received view in philosophy of consciousness today to consider the Hard problem and the explanatory gap problem to be one and the same problem. The notion of explanatory gap was introduced by J. Levine (1983). Chalmers (1995) developed Levine's idea of an explanatory gap into the Hard problem. In this article I show that there are two different problems here and give reasons why we should care about the explanatory gap rather than the Hard problem. Both Levine and Chalmers accept functional model of reductive explanation. According to them, functional reduction, that works so well for almost all phenomena, does not work in explaining consciousness. No cognitive function can grasp phenomenal consciousness. However functional model admits a compromise: a possibility to replace our pre-theoretical notion of consciousness with its theoretical successor-concept. I think that phenomenal representation is the best candidate for such successor-concept.

**Key words:** Consciousness, Hard problem, explanatory gap, functional reduction, replacement, successorconcept, phenomenal representation.

Представление о разрыве в объяснении сознания, как и понятие трудной проблемы, предполагают модель редуктивного объяснения, которая основана на концептуальном анализе и которую иногда называют «план Канберры». Особую роль концептуальному анализу в обосновании материализма придавали австралийские материалисты, а также американский философ Дэвид





## ПОЧЕМУ НАС ДОЛЖЕН ЗАБОТИТЬ РАЗРЫВ...

Льюис. Модель редукции была изложена Льюисом в виде аргумента [Lewis, 1972], который приводится ниже в изложении австралийского философа Фрэнка Джексона.

«1. Психическое состояние  $P$  = выполняет роль  $P$  (согласно концептуальному анализу).

2. Выполняющий роль  $P$  = состояние мозга  $M$  (согласно науке).

3. Следовательно,  $P$  = нервное состояние  $M$  (в силу транзитивности)» [Jackson, 1994].

И Левин, и Чалмерс принимают эту функциональную модель редуктивного объяснения. Левин описывает процесс редукции как двухступенчатый процесс: «Первая стадия включает (относительно? квази?) априорный процесс придания понятию редуцируемого свойства “формы” для редукции путем идентификации той каузальной роли, для которой мы ищем лежащие в ее основе механизмы [концептуальный анализ]. Вторая стадия включает эмпирическую работу по обнаружению того, каковы эти механизмы» [Levine, 1993: 132].

Чалмерс дает очень похожее описание редуктивного объяснения: «Для наиболее интересных феноменов, которые требуют объяснения, включая такие феномены, как репродукция и обучение, соответствующие понятия обычно можно проанализировать *функционально*. Суть таких понятий можно охарактеризовать в терминах выполнения некоторой функции или функций... Объяснив, как выполняются эти функции, мы объяснили данный феномен. Объяснив, как организм выполняет функцию производства другого организма, мы объяснили репродукцию, поскольку все, что означает репродукция, – это выполнение данной функции» [Chalmers, 1996: 43].

Чалмерс подчеркивает, что возможность концептуального анализа лежит в основе редуктивного объяснения. «Без такого анализа не было бы объяснительного моста от физических фактов более низкого уровня к объясняемому явлению. Имея же такой анализ, все, что мы должны сделать, – это показать, каким образом определенные физические механизмы более низкого уровня могут удовлетворять этому анализу, результатом будет объяснение» [Chalmers, 1996: 44].

Феномен считается редуктивно объясненным, если экспланандум *логически вытекает* из эксплананса. Левин пишет: «Редукция должна объяснить то, что редуцировано; и способ, с помощью которого мы можем узнать, достигнуто ли это, состоит в том, чтобы увидеть, делают ли редуцирующие феномены редуцируемый феномен эпистемологически необходимым, т.е. можем ли мы видеть, почему при наличии фактов, указанных в редукции, вещи должны быть именно такими, какими кажутся на первый взгляд [Levine, 1993: 129]. Чалмерс выражает ту же идею следующим образом: «Естественный феномен редуктивно объяснен в терминах свойств более низкого уровня именно



тогда, когда он логически проистекает (supervenient) из этих свойств» [Chalmers, 1996: 47–48].

Логическое (концептуальное) «проистечение» (supervenience) Чалмерс определяет так: «В-свойства логически проистекают из А-свойств, если не существует двух логически возможных ситуаций, которые идентичны в отношении их А-свойств, но отличаются в отношении их В-свойств» [Ibid.: 35]. Чалмерс подчеркивает, что логическое проистечение определяется не в терминах строгой дедуцируемости в какой-либо системе формальной логики, а более широко – в терминах возможных миров.

Согласно Левину и Чалмерсу, эта функциональная модель, которая так хорошо работает в объяснении почти всех феноменов, не работает при объяснении сознания. В случае сознания заключение логически не вытекает из посылок, т.е. существует разрыв в объяснении. Дело в том, что каузально/функциональный анализ понятия сознания упускает его квалитативный и субъективный аспекты. Левин вслед за Солом Крипке [Kripke, 1980] приводит пример с болью (этот стандартный пример устарел в свете данных современной физиологии, он выполняет лишь функцию иллюстрации идеи).

«В нашем понятии боли есть нечто большее, нежели ее каузальная роль, существует квалитативный характер боли, как она ощущается; и то, что открытие С-волокон оставляет без объяснения, – это почему боль должна ощущаться именно так, как она ощущается! Ведь в возбуждении С-волокон, кажется, нет ничего такого, что естественным образом соответствовало феноменальным свойствам боли по сравнению с некоторыми другими феноменальными свойствами» [Levine, 1983: 357].

Для того чтобы продемонстрировать, что в функциональном анализе сознания действительно что-то упущено, Левин и Чалмерс обращаются к *мыслимости* такой ситуации, когда функциональное описание сознания имеет место в действительности, но соответствующие квалиа отсутствуют (аргумент отсутствующего квалиа). Например, мыслимо, что возбуждение С-волокон имеет место, а ощущения боли нет. Таким образом, из того, что мы объяснили, как выполняется определенная функция, вовсе не вытекает, что мы объяснили сознание. Вот почему возникают дальнейшие вопросы: почему боль должна ощущаться именно так, как она ощущается? Почему вообще выполнение ее каузальной роли должно сопровождаться каким-либо переживанием?

И Левин, и Чалмерс полагают, что *любой* функциональный анализ сознания будет упускать квалитативный и субъективный аспекты сознания. «В той же мере, в какой в нашем понятии квалитативного характера остается элемент, который не схвачен свойствами его каузальной роли, он будет ускользать из объяснительной сети физикали-



## ПОЧЕМУ НАС ДОЛЖЕН ЗАБОТИТЬ РАЗРЫВ...

стской редукции» [Levine, 1993: 134]. Чалмерс пишет, что в объяснении существует разрыв между *функциями* и *переживаниями* и нам необходим объяснительный мост, чтобы преодолеть его. Простое объяснение функций находится на одной стороне этого разрыва, поэтому, по мнению Чалмерса, материалы для моста следует искать где-то в другом месте [Chalmers, 1995].

Если никакой функциональный анализ не может схватить качественный и субъективный аспекты сознания, то такой принципиальный разрыв в объяснении оказывается трудной проблемой в сравнении со всеми другими редукциями. Поскольку в модели функциональной редукции большое значение имеет концептуальный анализ, рассмотрим, как Чалмерс анализирует понятие сознания в своей книге «Сознающий ум» (1996).

Вначале Чалмерс различает феноменальный и психологический аспекты психики (mind): «Феноменальные понятия имеют дело с психическими аспектами от первого лица, а психологические понятия имеют дело с аспектами от третьего лица... Ни феноменальное, ни психологическое не должны быть определены в терминах друг друга» [Chalmers, 1996: 16]. Далее, вслед за Недом Блоком (1995), Чалмерс различает два понятия сознания: «феноменальное сознание» и «психологическое сознание», или «осознание» (awareness)<sup>1</sup>. Последнее включает в себя все или почти все различные психологические (когнитивные) понятия сознания; «быть феноменальным для состояния означает, что оно ощущается определенным образом, а быть психологическим для состояния означает, что оно играет соответствующую каузальную роль» [Ibid.: 12]<sup>2</sup>.

Два понятия сознания порождают две разные проблемы – «трудную» и «легкую». «Легкие проблемы сознания – это те, которые кажутся легко поддающимися стандартным методам когнитивной науки; с их помощью феномен объясняется в терминах вычислительных или нейронных механизмов. Трудная проблема не поддается этим методам» [Chalmers, 1995: 200]. «По-настоящему трудная проблема – это проблема переживания (experience)» [Ibid.: 201]<sup>3</sup>.

Модель функциональной редукции, по мнению Чалмерса, вполне подходит для решения легких проблем. «Объяснить способность к отчету, например, есть не что иное, как объяснить, каким образом система могла бы выполнить функцию создания отчетов о внутренних состояниях. Чтобы объяснить внутренний доступ, мы должны

<sup>1</sup> Последнее включает в себя все или почти все различные психологические (когнитивные) понятия сознания. «Быть феноменальным для состояния означает, что оно ощущается определенным образом, а быть психологическим для состояния означает, что оно играет соответствующую каузальную роль» [Chalmers, 1996: 12].

<sup>2</sup> Блок называет их «феноменальное сознание» и «сознание-доступность (access)» [Block, 1995].

<sup>3</sup> Под «переживанием» Чалмерс имеет в виду феноменальное сознание.



объяснить, как внутренние состояния могли бы надлежащим образом влиять на систему и как система могла бы использовать информацию об этих состояниях в контроле более поздних состояний. Чтобы объяснить интеграцию и контроль, мы должны объяснить, как центральные процессы системы могут объединять содержания информации и использовать их для обеспечения разнообразного поведения. Все это проблемы объяснения функций.

Как мы объясняем выполнение функции? Точно устанавливая *механизм*, который выполняет эту функцию. Самое лучшее для этой задачи – создать нейрофизиологические и когнитивные модели. Если мы хотим получить детальное объяснение на более низком уровне, можно установить нервный механизм, отвечающий за эту функцию. Если мы хотим более абстрактного объяснения, можно установить механизм в вычислительных терминах. В обоих случаях мы получим полное и удовлетворительное объяснение. Как только мы установили нервный или вычислительный механизм, который выполняет, например, функцию вербального отчета, большая часть работы по объяснению способности к отчету закончена... Это *концептуальный* факт об этих явлениях, что их объяснение включает лишь объяснение различных функций, поскольку эти явления *можно определить функционально*» [Chalmers, 1996: 4].

Однако, по мнению Чалмерса, феноменальное сознание нельзя определить функционально: «Что делает трудную проблему трудной и едва не уникальной, так это то, что она выходит за пределы проблем выполнения функций. Чтобы это увидеть, обратите внимание на то, что даже когда мы объяснили выполнение всех когнитивных или поведенческих функций в окрестности сознания – перцептуальная дискриминация, категоризация, внутренний доступ, вербальный отчет – все равно остаются без ответа дальнейшие вопросы: *Почему выполнение этих функций сопровождается переживанием?*.. Почему вся эта обработка информации не происходит “в темноте”, свободно от какого-либо внутреннего чувства?» [Ibid].

Я привела эту длинную цитату для того, чтобы продемонстрировать, что в результате концептуального анализа, данного Чалмерсом, феноменальное сознание *по определению* не может быть объяснено в функциональных терминах. Оно априори понимается как нечто нефункциональное, не поддающееся стандартным методам когнитивной науки. Чалмерс делает из этого метафизический вывод: сознание не редуцируемо к физическому, следовательно, оно не материально.

Критикуя антифункционализм и аргумент отсутствующего квалиа, на который он опирается, Роберт ван Гулик замечает, что эта линия аргументации допускает ошибку предвосхищения основания (question begging). «Как может антифункционалист знать априори, что любая функционалистская теория в будущем допустит реализа-



цию [функционального описания] без квалиа?» [Van Gulick, 1993: 146]. Ответ только один: в основе уверенности критиков функционализма в отношении сознания лежит произвольное разделение сознания на два – феноменальное и функциональное.

Разрыв в объяснении и трудная проблема, как было сказано выше, многим представляются одной и той же проблемой. Но мне кажется важным провести различие между ними. Начну с того, что в объяснении сознания сегодня действительно существует разрыв, т.е. существующим теориям сознания не удастся логически вывести качественный и субъективный характер сознания. Возможны две интерпретации этого разрыва: 1) он возникает, потому что в *случае* той или иной конкретной теории функциональный анализ сознания упускает его качественный и субъективный характер; 2) это принципиальный разрыв, потому что функциональный анализ качественного и субъективного характера феноменального сознания в *принципе* невозможен. Первая интерпретация вызывает вопрос: каким должен быть функциональный анализ, чтобы не упускать качественный и субъективный характер сознания? Это и есть проблема разрыва в объяснении. Вторая интерпретация ставит вопрос: как иначе объяснить феноменальное сознание, если оно не поддается редукции к физическим свойствам? Это трудная проблема сознания.

Иначе говоря, вопросы имеют разные предпосылки: предпосылкой первого вопроса является утверждение, что редуктивное объяснение сознания возможно, предпосылкой второго вопроса является отрицание этой возможности. На мой взгляд, трудная проблема сознания основана на произвольном концептуальном анализе, заранее исключающем возможность редуктивного объяснения.

Сегодня существует огромное число эмпирических свидетельств в пользу того, что между физическими процессами в мозге и сознанием существуют значительные корреляции. Это весьма правдоподобно говорит о том, что необходимая связь между ними есть. Но для того чтобы показать эту связь, нужен адекватный теоретический анализ сознания.

Каким образом может быть закрыт существующий разрыв? Очевидная трудность заключается в том, что наш экспланандум является обыденным понятием сознания. Это понятие включает в себя только субъективные переживания, данные от первого лица, и ничего более, ничего такого, что связывало бы сознание с мозгом. А для редукции к физическим процессам необходим анализ сознания в понятиях от третьего лица, ибо только так можно связать сознание с его физической реализацией.

Антиматериалистические аргументы (аргумент знания, аргумент мыслимости, аргумент разрыва в объяснении и их разновидности)



убедительно доказывают, что мы не можем вывести дотеоретическое понятие сознания из соответствующего редуцирующего описания. Мэри действительно узнает нечто новое: каково это видеть красный цвет. Нам действительно ничто не мешает помыслить зомби – существо, которое является физической копией человека, но не имеет внутреннего мира и для которого отсутствует *каково это быть* зомби. Но именно этого мы и должны были бы ожидать, принимая во внимание субъективный, от первого лица характер нашего дотеоретического понятия сознания. Значит ли это, что Чалмерс был прав, утверждая, что трудная проблема сознания – объяснения *каково это* – не поддается методам когнитивной науки и что даже когда мы будем знать все о когнитивных функциях и нейрофизиологии мозга, это не решит трудную проблему сознания? Думаю, что нет.

Дело в том, что план Канберры допускает компромисс. Физикализм кажется несовместимым с интуитивно правдоподобными высказываниями о природе вещей, которые кажутся нам существующими, таких, например, как субъективная реальность, личность, цвета, ценности, свобода воли и т.д. Как эта непоследовательность может быть устранена? Сторонники плана Канберры в качестве решения предлагают компромисс: надо ввести понятия-замены. Последние можно интерпретировать как понятия-преемники по отношению к интуитивным или обыденным понятиям, которые вызывают проблему. Понятия-замены позволят вписать спорные вещи в физическую картину мира.

Если бы у нас было адекватное понятие-замена понятию *каково это видеть* красный цвет, то мы бы не сомневались, что Мэри смогла бы вывести из имеющихся у нее знаний, каким образом человеческий организм способен ощущать красный цвет. Но это было бы, конечно, только теоретическое объяснение – с позиции третьего лица, но не первого.

В истории науки мы видим немало примеров того, как новые понятия заменяют старые в научной картине мира. Колин Макгинн приводит следующие аналогии: понятие веса и понятие массы, геоцентрическое и релятивистское понятия движения. Старое понятие трансформировалось в новое по мере созревания нашей теории мира. Новое понятие сохраняет ядро дотеоретического понятия, вместе с тем углубляя его и устраняя ошибочные ассоциации [McGinn, 2001: 304]. В случае сознания понятие-замена должно не только сохранить ядро дотеоретического понятия, но и сделать это в терминах теории от третьего лица.

Однако замена обыденного понятия новым чревата тем, что можно упустить феномен, который мы хотим объяснить, существует опасность подмены экспланандума. Об этом предупреждает Чалмерс: «Самый легкий путь разработать “теорию” сознания... это переопределить объясняемый феномен как нечто, чем он не яв-



ляется» [Chalmers, 1996: XII]. Чалмерс отмечает, что такой прием довольно часто встречается в теориях сознания. Переформулировав проблему как проблему объяснения определенных функций, превратив тем самым «трудную» проблему сознания в «легкую», можно создать элегантную теорию, но трудная проблема останется, уверен Чалмерс.

Чтобы избежать подмены, чтобы обеспечить, что понятие-замена есть понятие той же самой вещи, надо сохранить ядро нашего обычного понятия сознания – его качественный и субъективный характер. Понятие-преемник должно включать в себя эти свойства. Хорошим кандидатом на роль понятия-преемника, на мой взгляд, является понятие феноменальной репрезентации. Под феноменальной репрезентацией имеется в виду психическая репрезентация, средствами которой являются квалиа.

Репрезентационализм в отношении сознания поддерживает и Чалмерс. Он считает, что в пользу репрезентационализма существуют хорошие аргументы, но они не являются хорошими аргументами в пользу редуктивного репрезентационализма. Репрезентационализм отождествляет феноменальные свойства со свойствами репрезентации определенного содержания определенным способом. *Редуктивный* репрезентационализм утверждает, что эти репрезентациональные свойства можно полностью характеризовать в нефеноменальных терминах. *Нередуктивный* репрезентационализм утверждает, что невозможно объяснить в нефеноменальных терминах ни содержание, ни способ сознательной репрезентации. Таким образом, по мнению Чалмерса, репрезентационализм сам по себе не решает проблемы устранения разрыва [Chalmers, 2004].

В самом деле, можно ли проанализировать феноменальный способ репрезентации (репрезентации с помощью квалиа) в нефеноменальных терминах? Когда Фрэнк Джексон изменил свое мнение и вернулся на позиции физикализма, в качестве понятия-замены он принял понятие репрезентации. Наши феноменальные состояния, по его мнению, суть состояния репрезентации вещей. А психические репрезентации можно объяснить физически. Видеть красный цвет есть определенное репрезентациональное состояние. Разумеется, это очень сложная задача – найти такой анализ феноменальной репрезентации, который устранил разрыв между сознанием и физическими процессами в мозге. Джексон делает попытку выделить свойства, которые отличают феноменальную репрезентацию от нефеноменальной.

Он выделяет пять особенностей, отличающих феноменальную репрезентацию от нефеноменальной.

1. Феноменальная репрезентация отличается богатством. Визуальное восприятие представляет, каковы вещи здесь и сейчас, в терминах цвета, формы, локализации, протяженности, ориентации, движе-



ния. Тактильное восприятие представляет, каковы вещи, в терминах формы, движения, текстуры, протяженности, ориентации, температуры.

2. Она характеризуется неразрывностью: представление цвета одновременно представляет форму и локализацию.
3. Она непосредственна.
4. Она представляет вещи как воздействующие на нас.
5. Она играет особую функциональную роль в работе психики.

Если репрезентациональное состояние имеет все эти пять характеристик, то это состояние, по мнению Джексона, есть феноменальное [Jackson, 2004].

Можно ли принять это описание в качестве удовлетворительного теоретического анализа феноменальной репрезентации? Нет, потому что анализ дан в феноменальных терминах. Чтобы устранить разрыв в объяснении сознания, необходим такой теоретический анализ сознания, который бы определил в вычислительных терминах и квалиа, и субъективность сознания. Без этого разрыв в объяснении сознания не будет устранен.

## Библиографический список

- Block, 1995 – *Block N.* On a Confusion about a Function of Consciousness // Behavioral and Brain Sciences, 1995. Vol. 18. P. 227–247.
- Chalmers, 1995 – *Chalmers D.* Facing up to the Problem of Consciousness // Journal of Consciousness Studies. 1995. № 2. P. 200–219.
- Chalmers, 1996 – *Chalmers D.J.* The Conscious Mind // Search of a Fundamental Theory. Oxford : Oxford University Press, 1996.
- Chalmers, 2004 – *Chalmers D.J.* The Representational Character of Experience // B. Leiter (ed.). The Future for Philosophy. Oxford : Clarendon, 2004. P. 153–181.
- Jackson, 1994 – *Jackson F.* Finding the Mind in the Natural World // D. Chalmers (ed.). Philosophy of Mind: Classical and Contemporary Readings. Oxford University Press, 2002. P. 162.
- Jackson, 2004 – *Jackson F.* Mind and Illusion // P. Ludlow, Y. Nagasawa, D. Stoljar (eds). There's Something about Mary. Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument. Cambridge, Mass.; L. : The MIT Press, 2004. P. 421–443
- Kripke, 1980 – *Kripke S.* Naming and Necessity. Harvard : Harvard University Press, 1980.
- Levine, 1983 – *Levine J.* Materialism and Qualia: The Explanatory Gap // Pacific Philosophical Quarterly. 1983. Vol. 64. P. 354–361.
- Levine, 1993 – *Levine J.* On Leaving out What it is Like // M. Davies and G. Humphreys (eds). Consciousness: Psychological and Philosophical Essays. Oxford : Blackwell, 1993.





Lewis, 1972 – *Lewis D.* Psychophysical and Theoretical Identifications // Australian Journal of Philosophy. 1972. № 50. P. 249.

McGinn, 2001 – *McGinn C.* How Not to Solve the Mind-Body Problem // C. Gilet, B. Loewer (eds). Physicalism and its discontents. Cambridge : Cambridge University Press, 2001.

Van Gulick, 1993 – *Gulick van R.* Understanding the Phenomenal Mind: Are We All Just Armadillos? // M. Davies and G. Humphreys (eds.). Consciousness: Philosophical and Psychological Aspects. Oxford : Blackwell, 1993. P. 146.



## ЧТО ЗНАЧИТ БЫТЬ ТРАНСЦЕНДЕНТАЛИСТОМ В СОВРЕМЕННОЙ АНАЛИТИЧЕСКОЙ ФИЛОСОФИИ СОЗНАНИЯ?

**Диана Эдиковна Гаспарян** – кандидат философских наук, доцент, Национальный исследовательский университет Высшая школа экономики, факультет философии, кафедре онтологии, логики и теории познания.  
E-mail:  
anaid6@yandex.ru

В данном исследовании<sup>1</sup> я бы хотела показать, как работает трансцендентализм в рамках современной аналитической философии сознания и в каком смысле он противостоит натурализму. Далее я раскрою примеры трансценденталистских подходов в современной аналитической философии, а именно: теории К. Макгинна, Дж. Левина и Н. Хомского. Кроме того, я рассмотрю соотношение критических и позитивных компонентов в этих подходах, а также роль скептической позиции в их теориях.

**Ключевые слова:** трансцендентализм, аналитическая философия сознания, скептицизм.

## WHAT IS IT LIKE TO BE A TRANSCENDENTALIST IN THE CONTEMPORARY ANALYTIC PHILOSOPHY OF MIND?

**Diana Gasparyan** – candidate of philosophical sciences, Associate professor, National Research University Higher School of Economics, Department of Philosophy.

In this study I would like to demonstrate what is like to be a transcendentalist in contemporary analytic philosophy of mind. At the same time I will explore in what sense transcendentalism is opposed to naturalism. Next, I will consider the examples of transcendentalist theories as the approaches in contemporary analytic philosophy, namely the theory of K. McGinn, J. Levin and N. Chomsky. In addition, I will show, what is the proportion the negative and positive components in these approaches, and will stipulate the role of the skeptical position in their theories.

**Key words:** transcendentalism, analytical philosophy of mind, skepticism.



### Введение: неясность трансцендентализма

Термины «трансцендентализм» или «трансценденталистский подход» не слишком распространены в аналитической философии

<sup>1</sup> Хотелось бы особо поблагодарить своего коллегу Сергея Левина, без помощи которого этому исследованию, вероятнее всего, не суждено было бы появиться. Именно он инициировал его как некий развернутый ответ, спросив меня однажды, что означает высказывание «я трансценденталист» в ответ на вопрос о том, какая из существующих ныне теорий решения mind-body problem мне ближе.



сознания. Чаще всего они ассоциируется с философией Канта, включая всю совокупность ключевых положений его системы, таких, как априоризм, деление мира на феномены и ноумены, разделение на аналитические и синтетические суждения и т.д. Между тем можно говорить о трансценденталистском подходе, который был бы некоторым образом дистиллирован из кантовской философии и даже вообще представлял собой автономную философскую стратегию, к которой в разное время прибегали разные философы<sup>2</sup>. В этом случае его нужно понимать как определенный тип аргументации и чуть шире как определенный методологический принцип. Разумеется, и сам термин «трансцендентализм» мог бы быть произвольно заменен на любой другой, что не отменяло бы самой сути и внутренней логики этого подхода. Несмотря на скромную представленность в аналитических контекстах, трансцендентализм представляет собой серьезную альтернативу большинству имеющихся на сегодняшний день подходов к проблеме сознания. Кроме того, трансцендентализм является не просто еще одним подходом наряду с прочими, но скорее должен быть определен как метапрограмма, представляющая собой противовес другой большой метапрограмме натурализма [Smith, Sullivan, 2011].

Вместе с тем значительная часть аналитических философов не рассматривает трансцендентализм как «легальное» направление философии сознания и чаще полагает его сугубо историко-архивным явлением. В этом случае самоопределение в качестве трансценденталиста в аналитической среде оказывается не всегда понятным. Например, если кто-то заявляет, что в отношении ключевой проблемы сознание–тело он является трансценденталистом, то практически всегда от него требуются пояснения, что имеется в виду. Иными словами, в отличие от таких теорий, как теория тождества, субстанциальный дуализм или панпсихизм, значение позиции трансцендентализма по умолчанию не ясно.

Ниже я постараюсь прояснить суть данного подхода в конкретном применении к ключевым вопросам аналитической философии сознания и дать такие его характеристики, которые позволили бы безошибочно узнавать почерк трансценденталиста даже в случае, если он именуется как-то иначе.

### Трансцендентализм vs натурализм

Ключом к пониманию трансцендентализма является его критичность по отношению к натурализму. Возможно, именно в связи с этим не всем философам сознания аналитической традиции, где так силен дух позитивизма и натурализма, если использовать более современ-

<sup>2</sup> Например, Сократ, Платон, Кант, Витгенштейн.



ный термин, до конца ясен смысл трансценденталистского подхода. Для того чтобы очертить суть расхождений между этими двумя подходами, попробуем для начала выяснить их главные опорные пункты.

Согласно одному из наиболее известных определений Канта, трансцендентальным является «познание, занимающееся не столько предметами, сколько видами нашего познания предметов, поскольку это познание должно быть возможным a priori» [Кант, 1994: 121]. Эта основополагающая формулировка фиксирует первое правило трансцендентализма: нельзя не принимать во внимание условия возможности того опыта, который мы имеем. Познание должно начинаться не с вещей, а с условий возможности данности этих вещей в том виде, в каком мы их застаем: именно эти условия являются правильным началом. Напротив, натурализм всегда начинает с вещей: их он считает подлинным началом, а вопрос об условиях либо не вполне понятным, либо несколько надуманным [Kim, 1988]. Речь идет о двух типах вопросов: 1) как возможно, что есть то, что есть? и 2) что в мире есть в принципе? Если трансцендентализм ориентируется на первый вопрос, то натурализм на второй. Можно сказать, что трансцендентализм начинает с того, что не дано, а натурализм с того, что дано. Такая формула может показаться несколько эзотерической, но ниже я разведу ее загадочность.

Отличия натурализма и трансцендентализма можно сформулировать в виде следующих программных положений. Согласно натурализму,

- ◇ реальность онтологически однородна. Это положение имеет в виду отсутствие удвоенной онтологии, подразумевающей разные типы существования, а именно исходит из того, что мир один<sup>3</sup>;
- ◇ есть только один пакет онтологических законов и правил. Это положение оговаривает единство универсальных законов мироздания. Например, если физике присущи детерминизм и наблюдаемость в перспективе от третьего лица, то для этих правил не может быть исключений;
- ◇ все феномены, которые существуют, можно описать на одном языке. Речь идет о том, что если для описания объектных феноменов выбирается принцип исчислимости или наблюдаемости, то он может и должен быть атрибутирован и ко всем прочим феноменам;
- ◇ универсальный язык, пригодный для описания всего существующего, – это объектный язык или язык перспективы от третьего лица

<sup>3</sup> Чаще всего подразумевают, что этот единый мир – мир природы, т.е. вещных, физических объектов. Но это не совсем так. Было бы неверно отождествлять натурализм и материализм. Строго говоря, может быть и идеалистический натурализм. Или может встречаться монистический натурализм (например, тождество сознания и материи в расселовском монизме события [Russell, 2009]). Возможен натуралистический дуализм, оговаривающий единство и универсализм психофизических законов, тем не менее не отменяющих нередуцируемости ментальных фактов к физическим [Чалмерс, 2013].



[Gasser, 2007]. Имеется в виду публичный или приватный тип идентификации объектов; для натурализма он всегда публичный, так как речь идет об экстерналистском доступе к наблюдаемости объектов;

- ◇ эпистемология в потенциале может рассчитывать на эпистемический абсолютизм. Согласно этому принципу, в конечном итоге можно знать все, так как познанию препятствуют лишь количественные границы, однако никаких структурных или системных ограничений нет: оно неизбежно прогрессирует, но в режиме отсроченности;
- ◇ все условия той реальности, которую мы наблюдаем, принадлежат самой этой реальности. Поэтому вся реальность потенциально может быть понята и описана в конечном числе непротиворечивых теорий.

Отличительные черты трансцендентализма:

- ◇ реальность онтологически неоднородна. Это положение имеет в виду наличие удвоенной онтологии, подразумевающей разные типы существования, а именно: исходит из того, что есть мир, который подлежит наблюдению (познанию и теоретизации), и мир, который делает наблюдение возможным, но сам наблюдению (познанию и теоретизации) не подлежит;
- ◇ есть как минимум два пакета онтологических законов и правил. Это положение оговаривает отсутствие универсальных законов мироздания. Например, если физике присущи детерминизм и наблюдаемость в перспективе от третьего лица, то при описании воли субъекта или его сознания эти правила работать не будут и придется вводить принцип спонтанности и перспективы от первого лица;
- ◇ нельзя на одном языке описать все феномены, которые есть. Речь идет о том, что если для описания объектных феноменов выбирается принцип исчислимости или наблюдаемости, то он может и должен быть атрибутирован и ко всем прочим феноменам. Нетождественность различных языков описания можно показать на логических примерах с привлечением аппарата модальностей *de re* и *de se*, где в модусе *de re* мы имеем дело с *истинностью* того, что утверждается в высказываниях, и с соответствующими *положениями дел*, описанными в высказываниях, т.е. с референциями, в модусе *de se* устанавливается связь между высказанными *утверждениями* и описанными *положениями дел* и соответствующими психологическими реакциями, действиями и переживаниями самого познающего, такими, как убеждения, желания и намерения, его эмоциями и действиями в реальном мире [O'Dea, 2002];
- ◇ два отличных друг от друга языка – это, к примеру, объектный язык или перспектива от третьего лица, с помощью которых мы описываем вещи внешнего мира, предполагающие открытую и всеобщую



- регистрируемость, и язык, реализуемый в перспективе от первого лица, предполагающий привилегированный приватный доступ или приватную идентификацию [Baker, 2014];
- ◇ эпистемологии следует помнить о неизбежном скептицизме в отношении ряда вопросов, от которого не следует пытаться избавиться. В соответствии с этим принципом, если некоторые теории претендуют на прояснение структурных или системных условий формирования определенной реальности, то они должны оказываться либо противоречивыми, либо неполными, но в конечном счете не смогут быть когерентно успешными. Поэтому ни о каком эпистемическом абсолютизме речи не идет: нельзя знать все, так как есть принципиально «слепые пятна», являющиеся результатом структурных или системных ограничений;
  - ◇ у той реальности, которую мы наблюдаем, есть условия, не принадлежащие наблюдаемой реальности. Именно поэтому о них нельзя составить предметного знания – теории. Но то, что не подлежит опередмечиванию (и по поводу чего нельзя иметь теории), – своего рода ближайшее (не вынесено куда-то вовне как некое трансцендентное), поскольку является структурным условием всего сущего и организует наблюдаемую реальность. Подобное условие (в терминах классической трансцендентальной философии «условие возможности») обязательно присутствует во всех мыслительных построениях, но не может быть схвачено объективно в качестве предмета, значения или общего понятия.

Если как-то резюмировать эти положения, то мы получим формулу, которую я приводила выше: натурализм полагает, что мир полностью исчерпывается тем, что дано, и фактически все, что дано, эпистемически доступно (если не актуально, то потенциально – в перспективе развивающегося знания) [Paríneau, 1993]. Напротив, трансцендентализм утверждает: мир устроен так, что данное нам всегда есть результат или следствие некоторых системных требований к организации реальности, которые сами этой реальности не принадлежат и найдены в ней быть не могут. То, посредством чего реальность организуется, не есть часть реальности. Вероятное возражение натуралиста по поводу того, как же быть с известными нам законами естествознания, трансценденталистом легко отводятся: речь идет, разумеется, не об этом уровне, а об уровне более фундаментальном. Скорее это математика или логика и, разумеется, метафизика.

Важно понимать, что это правило работает безотносительно к тому, кто или что выступает организатором существующей реальности. Если мир как-то задан, то фундаментальные структуры этой заданности не будут принадлежать миру по крайней мере на тех же условиях, на которых существует то, что задающими структурами порождается. Структурная изнанка мира не будет его частью и не сможет быть най-



## ЧТО ЗНАЧИТ БЫТЬ ТРАНСЦЕНДЕНТАЛИСТОМ...

дена внутри самого мира. При этом вовсе не обязательно связывать эту организующую активность с субъектом, как это традиционно делается в различных версиях философии субъекта. К примеру, трансцендентализм кантовского типа привяжет наличие системных требований формирования реальности к субъекту, поэтому, если использовать витгенштейнианскую терминологию, субъект не будет частью мира – нигде в мире философского субъекта нет. Но эта привязка вовсе не обязательна<sup>4</sup>. Так, в платоновской или той же витгенштейнианской философии системные требования не привязаны к случайному и потенциально исчезающему субъекту, они суть объективные условия мира, но так же не видны и не проявлены в нем. Таковы, например, ценности у Платона и логика у Витгенштейна. Они суть те структуры, через которые и в которых сбывается мир, но они не встречаются в мире так же, как прочие объекты или факты. В этом тезисе только и состоит главный принцип трансцендентализма, который может быть представлен в версии как удвоенной онтологии (кантовский трансцендентализм), так и однородной (платоновский трансцендентализм). Об этих структурах невозможно сформировать предметного знания, им нельзя дать определение (Платон) и о них нельзя сказать, но их можно показать (Витгенштейн).

Они не являются чем-то неизведанным или скрытым, а, напротив, даны непосредственно и хорошо нам известны; скорее они сугубо функциональны и операциональны, разрешают собой пользоваться, но не допускают предметного схватывания, объективации, которая позволила бы говорить о них в перспективе от третьего лица или сделать их универсально наблюдаемыми. Структуры, о которых идет речь, всегда даны в режиме «как», но не в режиме «что». В таком случае приходится говорить о двух уровнях описания реальности – объектном и необъектном (модальности *de re* и *de se* [O'Dea, 2002], онтологии от первого и от третьего лица [Baker, 2011], факты и ценности, предложения и законы логики [Витгенштейн, 1958]). Если натурализм признает только объектный язык, то трансцендентализм указывает на то, что для определенных сфер реальности этот язык окажется нерабочим или будет вести себя ненормально, образуя парадоксы<sup>5</sup>.

В связи с этой интуицией Витгенштейн приводил знаменитую метафору глаза, который не видит себя в процессе видения<sup>6</sup>. В дополнение к этой метафоре можно привести еще один простенький пример: киномеханик, который прокручивает ленту в кинотеатре, не может

<sup>4</sup> Вместе с тем если принимать во внимание, что сочетание таких свойств, как не-объективируемость и непосредственная данность, является критерием трансцендентальной структуры, то тот факт, что философский субъект напрямую подпадает под эти характеристики, позволяет посчитать его природу трансцендентальной, а именно: предположить, что мир индуцируем субъектными структурами, как это и предполагается в классической философии субъекта.

<sup>5</sup> Например, парадокс «спящей красавицы».



появиться в самом фильме на правах персонажа или участника. Равно как мы не должны видеть пленку, если хотим смотреть фильм<sup>7</sup>. Иными словами, правила, организующие мир, в самом мире не представлены.

## Трансценденталистский подход в отношении проблем сознания

Если говорить непосредственно о философии сознания, то термин «трансцендентализм» будет использоваться в трех главных значениях. Во-первых, в значении антиредукционизма, согласно которому сознание не может быть сведено к существующим в проявленном мире предметам и фактам, а именно: к физическим объектам и процессам. Во-вторых, в значении, называемом также феноменологическим, согласно которому о сознании нельзя говорить в терминах объективно наблюдаемых сущностей или в перспективе от третьего лица, так как само сознание объектом не является. Кроме того, нельзя строить теории сознания, отталкиваясь не от самого сознания. Например, невозможно ставить вопрос о природе сознания начиная с устройства мозга или связи мозга с сознанием. В-третьих, в значении определенного эпистемического скептицизма, который будет прямо указывать на то, что в отношении двух важнейших проблем сознания, а именно: проблемы связи мозга с сознанием и проблемы роли сознания в каузальности мира, вряд ли будут найдены непротиворечивые решения<sup>8</sup>.

Рассмотрим последовательно эти три значения. Антиредукционизм в первую очередь направлен против различного рода физикалистских стратегий объяснения сознания. Согласно физикализму, сознание представляет собой физический процесс и может быть полностью редуцировано к физической активности мозга как ближайшему материальному субстрату. Напротив, антифизикализм полагает, что

<sup>6</sup> «5.633. Где в мире можно заметить метафизический субъект? Вы говорите, что здесь дело обстоит точно так же, как с глазом и полем зрения. Но в действительности вы сами не видите глаза. И не из чего в поле зрения нельзя заключить, что оно видится глазом» [Витгенштейн, 1958: 67].

<sup>7</sup> В традиции философии имеются и другие яркие метафоры, выражающие эту же мысль. Это, к примеру, интуиция Платона о том, что все дано нам в свете, но сам свет не дан (миф о пещере) или изречение Шопенгауэра: «То, что все познает и никем не познается, есть субъект» [Шопенгауэр, 1985: 154].

<sup>8</sup> Один из наиболее авторитетных современных авторов В. Васильев в своей книге, как кажется, соглашается с тем, что вопрос «как мозг порождает сознание?» возможно так и не удастся прояснить радикально, и в его отношении уместен скорее эпистемический пессимизм, чем оптимизм. Но проблема активной каузации сознания, по мнению автора, относится к числу вполне решаемых задач, за которую сам Васильев и берется [Васильев, 2013].





сознание не может быть описано как физический процесс и потому должно быть объяснено нередуктивными способами. Важно понимать, что антифизикализм еще не является трансцендентализмом, а лишь указывает на недостаточность редуктивистских объяснений и потому является необходимым, но не достаточным условием для трансценденталистской позиции.

Антиредукционизм в свою очередь указывает на такие аспекты сознания, которые не позволяют его редуцировать. Для удобства их можно разделить на две группы – эпистемологические и онтологические аспекты. *Эпистемологические аспекты* это в первую очередь:

- ◇ обладание «квалиа». Допустим, я пробую вино на вкус или созерцаю «Подсолнухи» Ван Гога. Можно сколь угодно тщательно изучить мой мозг и прочие телесные реакции организма, но исследователь не сможет получить доступ к тому, как я переживаю вкус вина или удовольствие от знаменитой картины, мое качественное состояние не станет его собственным качественным состоянием, а ведь только в этом случае он смог бы заполучить его в качестве объекта. Таким образом, квалиа есть такой способ переживания, который неотделим от того, кто его переживает. И в этом смысле квалиа не вполне является объектом, предполагающим сохранение при отделении от порождающих его форм;
- ◇ особый субъективный эпистемический доступ. Поскольку при кажущейся интуитивной ясности понятия «субъективность» оно все-таки довольно неопределенно, прояснить его можно с помощью таких характеристик, как приватность, привилегированный доступ или перспектива от первого лица. Группа этих весьма удобных концептов призвана пояснить не всегда очевидную редукционистски ориентированным философам идею того, что о своих сознательных актах человек знает прямо и непосредственно. Например, я не могу сомневаться в том, что испытываю боль или вижу желтое пятно на стене. Любое сомнение в этом случае может означать новое состояние сознания, но не может отменить старое, ибо его истинность дана наглядно, только как факт осознания. Положение непосредственности можно пояснить еще проще: свои феноменальные состояния мы не дедуцируем и не индуцируем, мы вообще их ниоткуда не выводим, потому знание о них не есть результат некоей сложной интеллектуальной работы, не есть продукт вывода. Странным выглядело бы умозаключение вида: «Всякий укол сопровождается ощущением боли; я укололся, следовательно, я испытываю боль». Наши ощущения самодостаточны, их не надо доказывать. Кроме того, непосредственность моих переживаний не может стать объектом наблюдения из перспективы третьего лица. Только я могу наблюдать свою боль, ее нельзя наблюдать со стороны так, как можно наблюдать мое тело, которое эту боль испытывает.



*Онтологические аспекты* указывают на определенный способ существования сознания. Это:

- ◇ непространственная природа сознания. Традиционное определение сознания включает в себя все признаки нематериальности, а именно: отсутствие протяженности, а также прочих физических характеристик, таких, как плотность, масса, вкус, цвет и проч. Сознание не есть вещь, и оно не может быть определено через набор вещественных атрибутов;
- ◇ интенциональность сознания. Принцип интенциональности в данном случае позволяет обрести важный теоретический критерий различия в способах существования физических и ментальных явлений. Если всякое сознательное состояние (представление) всегда предметно, то его можно толковать как объект, который в своем «что» совпадает со своим «как». Под представлением здесь понимается не то, что представляется, но сам акт представления. Тождество «что» и «как» ментальных объектов позволяет отличить их от объектов физических. Последние рассматриваются прямо противоположным образом – их «как» и «что» отличаются друг от друга. Мы знаем, что физические вещи могут отличаться от того, как они нам даны. Более того, когда мы говорим так, то говорим о самих вещах. Я не считаю, что то, как я воспринимаю вкус апельсина, совпадает с его реальной химической формулой, но я полагаю, что то, как я воспринимаю вкус, и есть все то, что я реально воспринимаю.

Именно эти фундаментальные свойства сознания уже на уровне одних только определений<sup>9</sup> заставляют говорить о сознании как особой несводимой к физической реальности сущности. Кроме того, исходя из этих свойств сознание не подпадает под стандартные модели описания или объяснения, по меньшей мере принятые в науке [Alter, Walter, 2010]. Научное исследование возможно лишь в условиях объективной картины действительности, исключая субъективные «точки зрения». Однако поскольку объективность сознания состоит в том, чтобы быть субъективным, т.е. сущность сознания состоит в субъективном опыте субъекта, оно (сознание) неизбежно ускользает из поля зрения ученых [Baker, 2011]. Неудача будет связана с тем, что знание всех физических фактов, составляющих суть ментальных состояний чьего-либо сознания (квалиа), еще не позволит нам пережить этот опыт как свой собственный, в то время как именно этот опыт и является объектом изучения. Если квалиа и есть тот объект, к изучению которого мы стремимся, то нам нужно получить именно его, а не то,

<sup>9</sup> Разумеется, для развернутого обоснования несводимости требуются более весомые аргументы, но так как они хорошо описаны в соответствующей литературе, я лишь упомяну их: аргумент от приватности, аргумент от логической возможности зомби, аргумент от эпистемологической асимметрии, аргумент от перевернутого спектра, аргумент от знания/квалиа [Гаспарян, 2014].



что делает его возможным. При этом мы не можем подменять его той нейрофизической корреляцией, которая сопутствует этому опыту. Но нельзя не считаться с эпистемической асимметрией двух этих точек зрения: одно дело непосредственно испытывать вкус апельсина, совсем другое – слушать рассказ о том, каков он на вкус. И если согласиться с тем, что субъективный опыт будет систематически ускользать из теорий, призванных объяснить связь ментальных фактов с физическими фактами, то перспективы создания удовлетворительной теории всерьез усложняются.

Под еще одним значением трансцендентализма подразумевается критика проблематизации сознания с точки зрения объекта. Согласно этой критике, пытаться говорить о сознании бессмысленно, так как при этом мы будем пользоваться тем же самым сознанием. Других познавательных инструментов у нас нет. Предметом сознания является само сознание – идет ли речь об индивидуальном опыте сознания или о сознании «вообще». Трансцендентализм в своей феноменологической интерпретации выступает против такого способа исследования, когда мы считаем, что у нас есть сознание как некий объект изучения и есть некий загадочный метаязык, не совпадающий с сознанием, на котором мы можем поговорить о сознании-объекте.

Эту идею можно пояснить на простом примере. Допустим, у нас есть некий исследователь сознания, который намерен изучать сознание в лабораторных условиях. Допустим также, что он – противник интроспекции и предполагает изучать не свое сознание, а чужое. Предположим, ему требуется составить представление о том, что чувствует другой человек, когда слушает ноктюрн Шопена или разглядывает картину Малевича. Перед нашим исследователем не стоит задача дать нейрофизиологическую картину реакций мозга при восприятии получаемой информации. Когда он приступает к исследованию, его задача состоит в том, чтобы воспроизвести ментальный опыт своего реципиента. В каком-то смысле это означает, что исследователю нужно почувствовать то же, что чувствует испытуемый. Возможно ли это? Что при этом будет чувствовать исследователь? Как нам подсказывает здравый смысл, он будет чувствовать свои переживания, возникающие у него при попытке представить то, что чувствует реципиент. Иначе говоря, очевидно, что исследователь останется наедине со своим собственным сознанием, несмотря на все попытки стать другим сознанием. Этот пример достаточно убедителен для случая, когда мы хотим увидеть мир чужими глазами. Когда мы говорим о трудностях воспроизведения чужого индивидуального сознания с помощью своего собственного, большинство людей соглашается с тем, что это задача почти невыполнима. Однако когда мы создаем различного рода теории сознания, мы не всегда замечаем аналогичные трудности. Трансценденталистская установка в данном случае указывает на то,



что подобные трудности есть и в тех случаях, когда мы пытаемся создавать теории сознания вообще, т. е. по сути объяснять сознание с помощью несознания. Например, физикалистские теории сознания, редуцирующие сознание к работе мозга, как будто бы занимают позицию, превышающую ментальный опыт, и стремятся дать ему физическое объяснение. Речь идет о том, чтобы говорить на языке физики и на этом языке объяснять сознание. Но это невозможно, так как физик как исследователь наделен сознанием, а его физическая картина или теория имеет смысл только внутри понимающего этот смысл опыта. Поэтому поставить перед собой задачу говорить о сознании как об объекте означает сформулировать парадокс.

При таком толковании даже антифизикализм останется скорее натуралистической позицией, чем трансценденталистской [Bell, 1999]. Дело в том, что какой бы ответ не предлагала та или иная теория, например сознание сводится к физическим процессам или сознание к ним не сводится, с точки зрения трансцендентализма любой ответ логически безупречен, так как уже является следствием смысловой работы сознания. Эта работа предшествует самому ответу и потому крайне трудно отделить то «сознание, о котором говорим», от того «сознания, которое делает возможным сам разговор». Только из опыта самого сознания, а не опыта мозговой деятельности можно адресоваться к сознанию и спрашивать о его специфике. Равно зависимость или независимость сознания от тела может быть выражена лишь на языке сознания, а именно в виде осмысленного высказывания, предполагающего сознательное понимание. Например, если мы попробуем провести такой мысленный эксперимент, в ходе которого попробуем представить свое сознание несуществующим, для успешного выполнения этого эксперимента (равно как и для констатации его невозможности) нам понадобится наше собственное сознание. Точно так же при утверждении, что сознания нет, но есть лишь физические состояния мозга или другой материи, сознание необходимо для осмысленности такого утверждения.

Таким образом, в попытке создать теорию сознания можно обнаружить принципиальную невыполнимость его редукции. В итоге все теории, пытающиеся объяснить сознание через то, что им не является, упускают из виду, что тематизация того, что не есть сознание, может быть выполнена только самим сознанием. Есть сознание сознания и сознание несознания. Это обстоятельство серьезно тормозит нас в попытках создать сколько-нибудь традиционную теорию сознания. «Сознание как предмет изучения» – конструкция столь шаткая, что мы сразу понимаем безнадежность наших усилий. Если всякая исследовательская рефлексия над сознанием совпадает с самим сознанием, то очевидно, что продвинуться в таком исследовании у нас вряд ли получится.



Наконец, третьим значением трансценденталистского подхода в философии сознания будет указание на эпистемический скепсис в отношении ключевых проблем сознания. Каков, например, будет ответ трансценденталиста на ключевой вопрос философии сознания: можно ли вообще решить mind-body problem? Вспомним самого Канта: несмотря на то что мы не встречаем у него прямых рассуждений об этой проблематике, можно попробовать реконструировать его мысль. Как кажется, его позиция в этом вопросе была бы следующей: поскольку сама эта проблема поставлена на ошибочном (догматическом) языке, любые варианты ее решения окажутся тупиковыми. В кантовской терминологии речь идет о том, что решение mind-body problem лежит по ту сторону всякого возможного опыта, а это надо понимать так, что достигнуто оно быть не может. В каком-то смысле стандартная редакция mind-body problem должна была бы подпасть под статус классической антиномии, но против этого говорит то, что число решений данной проблемы может превышать два возможных, например физикализма и антифизикализма. Кроме того, в отличие от структуры антиномии, где оба решения, тетическое и антитетическое, выглядят равно доказательными, доказательная сила физикализма, как кажется, несколько уступает аргументам антифизикализма. Но даже если данную проблему нельзя напрямую подвести под антиномическую структуру, свидетельствующую, согласно Канту, о том, что мышление пытается покинуть собственные пределы и создать теорию собственных условий возможности, в кантовской системе mind-body problem очевидно подпадает под задачу, некритически поставленную традиционной метафизикой и не принимающую во внимание границы познавательных возможностей.

В рамках кантовского трансцендентализма познающий субъект мыслит по эту сторону возможного опыта и никак не может покинуть его актуальную почву [Кант, 1994]. Все, что мы можем знать, связано с тем, что дано нам в опыте, который образован посредством нашего действительного участия. Мы не получаем данные о мире в качестве нейтрального материала, но обрабатываем внешнюю информацию (чувственно и концептуально) и лишь затем приступаем к ней в качестве предмета нашего знания [Allison, 2004]. Для того чтобы выяснить, как устроен наш аппарат, мы должны были бы превысить тот порядок вещей, который дан нам в опыте, что невозможно. Это ограничение носит объективный характер, поскольку мы можем познать лишь то, чем не являемся сами. В свою очередь, так как наше сознание есть то единственное, с помощью чего мы пытаемся понять свое же собственное сознание, эта задача вряд ли когда-нибудь окажется реализованной.

Таким образом, если говорить о трансценденталистской позиции вообще, то проблемы, связанные с устройством сознания, будут свя-



заны с трудностями проблематизации системных требований, которые, как мы говорили выше, всегда должны оставаться в тени. Нет и не может быть непротиворечивой и полной теории, объясняющей устройство системных условий возможности. Индикатором того, что система упирается в собственные основания и пытается формализовать то, что служит условием возможности ее работы и определенного способа формализации, является появление противоречивых, неполных или антиномически устроенных теорий. Серьезным недостатком такого рассуждения будет радикальный эпистемический скепсис. Но если мы допустили, что у реальности есть условия, то избежать такого вывода будет крайне сложно. При трансценденталистском устройстве мира у нас всегда будут сохраняться такие вопросы, которые не могут быть решены [Stern, 1999]. Проблемы сознания относятся к числу именно таких вопросов. В терминологии Чалмерса они «трудны», а в терминологии трансценденталиста либо бессмысленны уже на уровне вопросов (Витгенштейн), либо все же, будучи осмысленно поставлены, не могут привести к когерентно выверенным решениям.

## Трансценденталисты в аналитической философии сознания

На сегодняшний день в рамках аналитической традиции можно назвать трех главных философов, чьи подходы и теории с некоторыми оговорками могут быть расценены как трансценденталистские. Не удивительно, что все они выглядят скептическими и предлагают скорее критическую программу решения ключевых проблем сознания, чем позитивную. Я бы отнесла к ним К. Макгинна, Н. Хомского и Дж. Леви, чьи теории с некоторыми оговорками можно посчитать трансцендентальными и в этом смысле скептическими, что я и попробую показать.

Наиболее близкой к скептицизму представляется теория когнитивной замкнутости, разработанная Колином Макгинном. Эта теория напрямую перекликается с философией Канта, у которого можно найти критические соображения по вопросу перспективы решения психофизической проблемы (восходящей к картезианской постановке проблемы, как сознание соотносится с телом). В кантовской терминологии речь идет о том, что решение проблемы как сознание связано с телом, лежит по ту сторону всякого возможного опыта, а это надо понимать так, что достигнуто оно быть не может. Данное положение перекликается с тем, что подразумевает Макгинн под понятием «когнитивная замкнутость». Тип мышления *M* когнитивно замкнут по отно-



шению к свойству  $P$  (или теории  $T$ ), если процедуры формирования понятий в рамках мышления  $M$  не могут быть распространены за его собственные пределы, чтобы освоить некое  $P$  (или понять теорию  $T$ ). Сознания могут быть разных типов и при этом они оснащены различными мощностями и ограничениями, предубеждениями и слепыми пятнами, так что определенные свойства (или теории) могут быть доступны одним сознаниям, но закрыты для других. То, что закрыто для сознания крысы, может быть открыто для ума обезьяны, и то, что открыто для нас, может быть закрыто для обезьяны. Сила познания не устроена по принципу «все или ничего», в каких-то эпистемических задачах мы успешны, а в каких-то нет. Сознание есть такой же биологический продукт, как организмы, и так же, как организмы, органы сознания могут быть различных форм и размеров и соответственно более или менее подходящими и релевантными для решения определенных когнитивных задач [McGinn, 1999: 67].

Как видно, Макгинн толкует ограничения познания в довольно натуралистическом ключе. Речь идет о том, что у человека есть некие биологические параметры (строение мозга или органов чувственности), которые выступают его системными ограничениями. В силу этого обстоятельства, несмотря на то что связь сознания с телом действительно является не надуманной проблемой, мы принципиально не способны дать на нее удовлетворительный ответ. Вопрос о природе сознания попросту лежит за пределами наших когнитивных способностей, поскольку каждый биологический вид имеет определенные ограничения. Например, собаки не в состоянии доказать теорему Пифагора, а люди не в состоянии создать теорию, объясняющую, как сознание связано с телом. При этом Макгинн не подразумевает субъектоцентристскую модель мира: речь не идет о том, что мир, как он нам дан, сформирован посредством априорных структур трансцендентального аппарата субъекта. Все дело лишь в том, что мыслительные способности человека ограничены рамками своего природного своеобразия. Иными словами, возможно, было бы поспешным считать позицию Макгинна по умолчанию трансценденталистской. Однако как бы то ни было, из такой модели нетрудно получить тезис об ограниченности наших познавательных способностей по крайней мере на том участке, который имеет отношение к устройству самого нашего познания. В связи с подобного рода скепсисом теорию Макгинна принято называть в аналитической философии сознания мистериальной [McGinn, 1999]. С точностью и адекватностью этого термина можно было бы поспорить, так как в скептицизме, строго говоря, нет ничего мистического. Если скептик аргументирует, в связи с чем не может быть получен ответ на определенный вопрос, его позиция будет выглядеть скорее реалистической, чем мистической. Напротив, попытки иметь позитивное знание по поводу тех областей, которые





в силу некоторых системных ограничений не могут стать предметом исследовательской процедуры, часто выглядят как мистицизм. Впрочем, удачность термина играет здесь сугубо второстепенную роль, нам интереснее скорее само отношение к перспективе решения проблемы сознания-тела, которое у Макгинна и его сторонников является скорее пессимистическим.

Еще одним философом-скептиком, который, пожалуй, не скрывает своей пессимистической позиции в отношении mind-body problem, является Ноам Хомский. Хомский предлагает поделить все задачи, за решение которых когда-либо брался человеческий рассудок, на загадки (тайны) и проблемы. Проблемы могут иметь решение, а загадки – нет. Все дело в том, что проблемы изначально составлены так, что представляют собой переформулированные и скрытые решения, и задача исследователя состоит только в том, чтобы расшифровать заложенный в вопросах ответ. Подобное положение не нужно расценивать так, что все задачи, которые может решить человеческий рассудок, являются скрытыми тавтологиями. Большая часть задач и решений лежит в области экспериментального знания, а львиная доля теорий, которые человечество создает, является частью эмпирической науки, а именно: естествознания. Когда мы говорим, что в случае проблем исследователю достаточно лишь расшифровать уже заложенный в вопросе ответ, то подразумеваем, что даже научные проблемы ставятся при участии интерпретаций и ожиданий, привносимых ученым. Тот материал, с которым ученый имеет дело, таков, что он релевантен аппарату и инструментам познавательных процедур.

Согласно Хомскому, необходимо различать допустимые и недопустимые гипотезы. Человеческий разум – это система свойств, образованная определенными полномочиями и ограничениями. Соответственно допустимые гипотезы – это те, которые доступны для определенной специфической системы, в данном случае системы человеческого разума. В отношении этих гипотез разум может построить достаточно богатые и сложные объяснительные теории. Но те же самые свойства разума, которые позволяют успешно выдвигать и проверять успешные гипотезы, могут исключать работу с другими недоступными гипотезами как иррелевантными человеческому разуму. Некоторых теорий может попросту не быть среди тех, которые определяются специфическими свойствами человеческого разума, притом что они могут быть доступны иначе организованному интеллекту [Chomsky, 1975: 121–124]. Вопреки натуралистической установке, согласно которой сознание человека является частью природного мира и, следовательно, имеет потенциал к решению любых естественно-научных проблем, имеются некоторые рамки и ограничения, которые определяются особой структурой самого сознания. Хомский прибегает к примерам, с которыми мы уже встречались у Макгинна:





для крысы некоторые вопросы являются тайнами, которые выходят за пределы ее познавательной деятельности, и потому она не может их решить. Подобная ситуация имеет место и в отношении человека. Те усилия, которые возлагаются на развитие «естественных наук», а именно: обнаружение свойств человеческого сознания как аспектов природного мира, могут никогда не увенчаться успехом, так как область сознания может оказаться как раз той областью, куда человеческий разум не сможет проникнуть в силу системных ограничений [Chomsky, 1975: 122–125].

Еще одним вариантом демонстрации трудностей позитивной теории сознания является подход Джозефа Левина. В первую очередь его критика направлена против физикализма и лишь во вторую ставит под вопрос перспективу получения адекватной теории сознания вообще. Физикализм проваливается потому, что должен не просто снабжать нас физическим описанием ментальных состояний и свойств, он должен давать объяснение этих свойств. Однако проблема в том, что такое «объяснение» может возникнуть не только в рамках физикализма, но и любой другой теории, стремящейся дать непротиворечивое решение mind-body problem. Левин приводит следующий пример: как можно объяснить тот факт, что Цицерон тождествен Туллию? Можно объяснить лишь, что один и тот же объект носит два имени. Левин же указывает, что в данном случае объяснение должно вовлекать дедуктивное отношение между экспланансом и экспланандумом. Некто может объяснить *P*, если он способен дедуцировать *P* из утверждений об экспланансе. Примером может служить объяснение феномена кипения воды, где появление пузырьков может быть объяснено ссылкой на межмолекулярное взаимодействие элементов воды.

Если два высказывания дедуктивно связаны друг с другом, то между ними имеется необходимая связь. Однако такой необходимой связи нет между высказываниями о мозге и феноменальном сознании. Между ментальными фактами и физическими неизбежно сохраняется некий провал в объяснении. Провал в объяснении – это термин, обозначающий трудность физикалистских теорий сознания в объяснении того, как физические свойства порождают способ ощущения вещей в момент их восприятия. К примеру, утверждение «боль есть стимуляция С-волокон» является информативным в физиологическом смысле, но никак не помогает нам понять, как переживается боль. Чтобы привести пример явления, в котором нет объяснительного провала, представим современный компьютер: есть что-то чудесное в том, что мы можем полностью объяснить его работу с помощью одного устройства его составных частей и наоборот. Напротив, субъективный сознательный опыт образует отдельный эффект, требующий отдельной причины, которую приходится или помещать вне фи-





зического мира или объявлять неизвестным физическим явлением (Levine, 1983: 87–89).

В общем виде обозначенные в этих трех подходах проблемы имеют вид метаязыкового парадокса. Мы пытаемся сделать объектом то, что должно являться также и инструментом, и в этом случае процедура объективации не может быть выполнена. Это во многом связано с тем, что мы пытаемся получить доступ к сознанию при помощи тех же логических инструментов, которые являются неотъемлемой частью фундаментальных свойств самого сознания. В этом случае, впрочем, не ясно, что может выступить своего рода инструментальным метаязыком, позволяющим провести непротиворечивую процедуру описания. Более того, сознание само по себе появляется только при условии работы самих этих инструментов (системных условий возможности). Но в этом случае сознание нельзя определить в терминах определяемого и определяющего, объясняемого и объясняющего, так как оно может или всегда оставаться только объясняющим, или совпадать и с тем, и с другим, что обесмысливает всю процедуру объяснения в целом.

### **Заключение: испытание скептицизмом**

Итак, самый важный тезис трансцендентализма состоит в том, что условия формирования мира сами в мире не встречаются. Это общее положение трансцендентализма напрямую срывает в отношении проблем сознания: сознание попросту не встречается в мире; мы точно знаем, что оно есть (в качестве функционального «как»), но никогда не найдем его посреди всех доступных предметов и фактов (в качестве некоего «что»). Строго говоря, такая формулировка предполагает скептицизм в качестве эпистемической позиции. Скептицизм становится серьезным испытанием трансцендентализма [Hookway, 1999]. К скептицизму в философии всегда относились скептически: в случаях, когда он был следствием своего рода интеллектуального кокетства, его рассматривали как позерство, а в случаях, когда он заявлял о себе вполне серьезно, его считали формой малодушной капитуляции. Как бы то ни было, философу всегда существенно сложнее, чем ученому, признаться в том, что он не знает, как решать ту или иную проблему. Это связано с традиционными различиями в методах научного и философского исследования. Если наука ориентирована на поиск решений во времени, выдвижение и отбрасывание гипотез, то философия надеется обрести универсальное решение, которое, будучи вечной истиной, было бы неподвластно времени. Такое представление о философских способах достижения истины всерьез пере-



смачивается в современной философии, особенно аналитической, и в наши дни философы все чаще заявляют о том, что и в философских рефлексиях завтрашний день может оказаться эффективнее сегодняшнего. Но, несмотря на проникновение в философию ожиданий, свойственных науке, возможность занять скептическую позицию по-прежнему сопровождается у философов серьезным дискомфортом: скептические версии оставляют нас наедине с нашими догадками и скорее отговаривают от дальнейшего поиска. В свое оправдание скептики часто заявляют, что не просто бойкотируют исследование, а помогают понять и осознать, почему окончательное решение той или иной проблемы оказывается принципиально недостижимым. Скептицизм дает определенную ясность, указывая истинные причины тех затруднений, с которыми сталкиваются философские теории [Skorupski, 1990].

В итоге мы получаем следующую дилемму. Натурализм предлагает позитивное знание, но, будучи абсолютизированным, он либо обольщается ложной картиной (в самом начале предполагая редуцированную вселенную), либо строит противоречивые теории (например, физикализм в виде теории тождества). Трансцендентализм же прокладывает границы, но прокладывает их правильно: он хотя и не индуцирует позитивное знание, но и не образует противоречий. При этом может возникнуть опасение, что трансцендентальное знание неполно, но это не совсем так: оно полно, так как затрагивает все элементы, но в отношении тех (структурных) компонентов, которые не предполагают опредмечивания, трансцендентализм высказывается в критическом духе невозможности создания законченной теории [Franks, 2005].

Учитывая сказанное, можно попытаться понять ситуацию, которая сложилась в области современной аналитической философии. Сохраняющаяся популярность натурализма объясняется тем, что у трансцендентализма нет содержательных альтернатив в виде объяснительных теорий, хотя и имеется серьезная и нередко почти разрушительная для натурализма критика [Cassam, 2009]. В связи с этим возникает ощущение, что философы пытаются держаться подходов хоть и не безупречных, но дающих некоторое позитивное знание, с которым можно либо соглашаться, либо нет [Kitcher, 1992: 68–69]. Так, в силу некоторой изначальной непопулярности скептических позиций указание на нерешаемость mind-body problem в области аналитической философии сознания выглядит для исследователей слишком большим разочарованием, чтобы этот подход мог претендовать на массовое признание. Несмотря на то что антинатурализм представлен далеко не только скептическими теориями, антинатуралистские программы можно охарактеризовать как в основном критические. Главный акцент в них сделан на критике существующих физикалист-



ских теорий и проблематичности, зачастую радикальной, ряда существующих решений mind-body problem. Таким образом, хотя трансцендентализм выступает сегодня с достаточно серьезной критикой натурализма, позиции последнего еще достаточно сильны. Изменится ли эта ситуация, напрямую зависит от того, насколько восприимчивым окажется натурализм к критическому проекту формирования знания.

### Библиографический список

- Васильев, 2013 – *Васильев В.В.* Сознание и вещи. М. : URSS, 2013
- Витгенштейн, 1958 – *Витгенштейн Л.* Логико-философский трактат. М., 1958.
- Гаспарян, 2014 – *Гаспарян Д.Э.* В защиту феноменального сознания: аргументы против физикализма в современной аналитической философии. Ч. 1 // Философские проблемы информационных технологий и киберпространства. 2014. № 1. С. 43–55.
- Гаспарян, 2014 – *Гаспарян Д.Э.* В защиту феноменального сознания: аргументы против физикализма в современной аналитической философии. Ч. 2 // Философские проблемы информационных технологий и киберпространства. 2014. № 2. С. 52–85.
- Кант, 1994 – *Кант И.* Критика чистого разума // И. Кант. Соч. Т. 2. М., 1994.
- Чалмерс, 2013 – *Чалмерс Д.* Сознание и ум: в поисках фундаментальной теории. М. : URSS, 2013.
- Шопенгауэр, 1992 – *Шопенгауэр А.* Собр. соч. В 5 т. М. : Московский клуб, 1992. Т. 1.
- Allison, 2004 – *Allison H.* Kant's Transcendental Idealism: An Interpretation and Defence. New Haven : Yale University Press, 2004.
- Alter, Walter, 2010 – *Alter T., Walter S.* Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism. Oxford : Oxford University Press, 2010.
- Baker, 2011 – *Baker L.R.* Does Naturalism Rest on a Mistake? // American Philosophical Quarterly. 2011. Vol. 48, № 2.
- Baker, 2014 – *Baker L.R.* Can Subjectivity be Naturalized? // Metodo: International Studies in Phenomenology and Philosophy (forthcoming).
- Bell, 1999 – *Bell D.* Transcendental Arguments and Non-Naturalistic Anti-Realism. Transcendental Arguments: Problems and Prospects. Oxford : Oxford University Press, 1999.
- Cassam, 2009 – *Cassam Q.* Can Transcendental Epistemology be Naturalised? // The Philosophical Review. 2009. Vol. 90.
- Chomsky, 1975 – *Chomsky N.* New Horizons in the Study of Language and Mind. Cambridge, 1975.
- Franks, 2005 – *Franks P.* All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism. Cambridge, MA: Harvard University Press, 2005.



Gasser, 2007 – *Gasser G.* Naturalism and the First-Person Perspective. How Successful is Naturalism? Frankfurt a/M : Ontos Verlag, 2007.

Hookway, 1999 – *Hookway C.* Modest Transcendental Arguments and Sceptical Doubts: A Reply to Stroud // *Stern*. 1999. Vol. 76.

Kim, 1988 – *Kim J.* What is «Naturalised Epistemology»? // *Philosophical Perspectives*. 1988. Vol. 2.

Kitcher, 1992 – *Kitcher P.* The Naturalists Return // *Philosophical Review*. 1992. Vol. 101.

Levine, 1983 – *Levine J.* Materialism and Qualia: the Explanatory Gap. Oxford : Oxford University Press, 1983.

McGinn, 1999 – *McGinn C.* Mysterious Flame: Conscious Minds in a Material World. N.Y., 1999.

O’Dea, 2002 – *O’Dea J.* The Indexical Nature of Sensory Concepts // *Philosophical Papers*. 2002. Vol. 31, № 2.

Papineau, 1993 – *Papineau D.* Philosophical Naturalism. Oxford : Blackwell, 1993.

Russell, 2009 – *Russell B.* An Outline of Philosophy. L. ; N.Y. (repr.), 2009.

Skorupski, 1990 – *Skorupski J.* The Intelligibility of Scepticism. Oxford : Basil Blackwell, 1990.

Smith, Sullivan, 2011 – *Smith J., Sullivan P.* Transcendental Philosophy and Naturalism. Oxford : Oxford University Press, 2011.

Stern, 1999 – *Stern R.* Transcendental Arguments and Scepticism: Answering the Question of Justification. Oxford : Clarendon Press, 1999.



## О ПРОВЕРГАЕТ ЛИ АРГУМЕНТ КАУЗАЛЬНЫХ ТРАЕКТОРИЙ ЛОКАЛЬНУЮ СУПЕРВЕНТНОСТЬ МЕНТАЛЬНОГО НАД ФИЗИЧЕСКИМ?

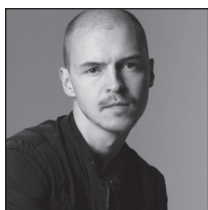
**Дмитрий Борисович Волков** – кандидат философских наук, содиректор НП «Московский центр исследования сознания». E-mail: [dvolkoff@gmail.com](mailto:dvolkoff@gmail.com)

Тезис о супервентности (ТС) ментального над физическим является общепринятым в современной философии сознания. Аргумент каузальных траекторий (АКТ1), предложенный В.В. Васильевым, направлен против этого тезиса, в частности против локальной супервентности. В настоящей статье автор защищает ТС и показывает, что если аргумент подвергает сомнению локальную супервентность, он скорее всего подвергает сомнению и глобальную супервентность. Это своеобразное *reduction ad absurdum* аргумента. Однако данный вывод не окончателен. Более надежным является второй способ опровержения аргумента – доказательство от противного. В статье предложен новый аргумент (АКТ2), сходный по своей структуре с аргументом каузальных траекторий, но приводящий к очевидно ложным выводам. Если аргументы АКТ1 и АКТ2 аналогичны, то оба аргумента ложны. В статье рассмотрено возможное возражение – указание на то, что АКТ1 и АКТ2 различаются из-за отсутствия в АКТ2 квалиа, показано, что и в АКТ1 квалиа не имеют значения. В заключение представлено решение парадокса двойников с разными биографическими историями. Парадокс предложен Васильевым в качестве иллюстрации абсурдности следствий ТС. Однако более подробное рассмотрение этого мысленного эксперимента показывает, что тезис о локальной супервентности не ведет к абсурдным следствиям. Таким образом, представляется, что ТС защищен.

**Ключевые слова:** психофизическая проблема, локальная супервентность, интеракционизм, аргумент каузальных траекторий.

## DOES THE ARGUMENT FROM CAUSAL TRAJECTORIES UNDERMINE LOCAL SUPERVENIENCE OF MENTAL ON PHYSICAL?

**Dmitry Volkov** – Ph.D., co-director «Moscow Center for Consciousness studies».



Thesis of supervenience (TS) of mental on physical is a common ground for many contemporary analytic philosophers. The argument from the causal trajectories presented by V. Vasilyev (ACT 1) is challenging this principle, in particular the local supervenience. In this paper the author defends the thesis of supervenience. He demonstrates that if the argument from causal trajectories undermines the local supervenience, it also undermines the global supervenience of mental on physical. This is a kind of *reductio ad absurdum* of the initial argument. However this may sound inconclusive. The author presents a more reliable way to disapprove the argument – proof by contradiction. In this paper he suggest another argument (ACT 2) which is analogous the ACT 1, but which leads to obviously false conclusions. If two arguments are analogous, then both arguments must be false. The author of this paper considers possible objections. The first objection has to do with qualia. One may suggest that two arguments are different because of the lack of qualia in the ACT 2. However it is demonstrated that qualia plays no significant role in the first argument as well. So it doesn't make a relevant difference. The second objection is the superficial absurdity of the consequences of TS suggested by Vasilyev, namely the thought experiment with the two identical persons with different auto-



biographical causal trajectories. This paradox is supposed to demonstrate the inconsistency of TS. However it is shown in the paper that there are no inconsistencies and the thought experiment doesn't demonstrate that the TS is false. Thus thesis for supervenience is defended.

**Key words:** supervenience, mind-body problem, interactionism, mental and physical properties, argument from causal trajectories.

## Отношение супервентности

Супервентность – это характеристика отношения между различными свойствами объекта. Набор свойств А является супервентным по отношению к набору свойств Б в том случае, если объект не может различаться в свойствах А без различий в свойствах Б<sup>1</sup>. Супервентность может быть локальной и глобальной. Если супервентность утверждается по отношению ко всему миру – это утверждение глобальной супервентности, а если в отношении индивидуальных объектов – локальной супервентности. Локальная супервентность – более сильное утверждение. Если в мире существует локальная супервентность, то в нем также присутствует и глобальная супервентность, но не наоборот.

Супервентными являются различные высокоуровневые свойства по отношению к низкоуровневым, фундаментальным. Например, биологические свойства супервентны над физическими, т.е. объект, идентичный по физическим свойствам, идентичен и по биологическим свойствам. Скажем, два животных, идентичные по своей физической структуре, не могут быть одновременно жирафом и нежиром. Над физическими свойствами супервентны также эстетические свойства. Одинаковые в физическом плане картины не могут различаться по красоте. Но одинаковые по своим физическим и эстетическим свойствам картины могут различаться по своей ценности. Представим, что существует идеальная копия картины Мунка «Крик». По своим эстетическим характеристикам она полностью соответствует оригиналу. Но ее ценность будет совсем иной. В произведениях искусства важен не только фактический результат, но и его оригинальность. Принадлежность кисти конкретного художника к исторической эпохе является ключевым при определении ценности произведения. Поэтому ценность произведения искусства не является локально супервентным над его физическими свойствами. Отношение между физическими свойствами и ценностью произведений искусств – это отношение глобальной супервентности.

Супервентность скорее всего существует между физическими и ментальными характеристиками. Ведь ментальные характеристики

<sup>1</sup> Одним из первых этот термин в современном смысле стал употреблять оксфордский философ Р. Хэйр [Hare, 1984].



зависимы от физических и изменения физических характеристик ведет к изменению ментальных. В этом смысле утверждение супервентности является частью решения психофизической проблемы. Распространение этого понятия в современной философии сознания связано с Дональдом Дэвидсоном. В статье «Ментальные события» (1970) он писал: «Ментальные события в некотором смысле зависимы или супервентны над физическими характеристиками. Под супервентностью я здесь подразумеваю невозможность существования двух событий, идентичных в физическом плане, но различающихся в ментальном, а также невозможность изменения ментальных характеристик без изменения физических» [Davidson, 2006: 214]. Вслед за Дэвидсоном многие философы-аналитики стали исследовать это отношение применительно к сознанию. Среди них Дж. Ким, Д. Льюис, Д. Чалмерс и др. [Kim, 1998; Kim, 2006; Kim, 2010; Lewis, 1983; Чалмерс, 2013]. Традиционным является утверждение локальной супервентности свойств сознания над мозгом. Такой позиции придерживается большинство философов. В этом контексте наиболее любопытной является попытка опровергнуть локальную супервентность сознания над мозгом с помощью аргумента каузальных траекторий современного российского философа В.В. Васильева.

### **Аргумент каузальных траекторий против локальной супервентности**

Аргумент Васильева сформулирован в виде трех посылок и вывода. Первая – это утверждение о том, что одинаковые события могут иметь разные причины. Вот шар попал в лузу, он мог там оказаться самыми различными способами, в результате самых различных соударений и движения по множеству траекторий. Как, впрочем, любая вещь могла оказаться там, где она есть, множеством различных способов. Посылка выглядит бесспорной. Вторая посылка – это констатация того, что память отражает события, происходившие с субъектом в прошлом, т.е. отражает каузальную историю субъекта. Конечно, память может представлять каузальную историю ошибочным образом, но это скорее исключение, чем норма. Столь же несомненно, по мнению Васильева, что воспоминания представляют собой приватные квалиативные состояния и образуют основу убеждений, желаний и других осознанных состояний субъекта. Третья посылка указывает на корреляцию поведения и желаний. Эта посылка не отрицается даже эпифеноменалистами, которые, правда, считают, что корреляция не имеет отношения к причинно-следственной цепи. А вот представленный аргумент заставляет нас склониться к обратному. Объединяя





посылки, Васильев получает следующую картину. «Если реально, что одно и то же событие или комплекс событий могли бы порождаться разными причинами и если мой мозг в данный момент времени может рассматриваться как громадный комплекс физических событий (некоторые составляющие которого могли бы иметь и иные причины), то, очевидно, что он мог оказаться в данном состоянии в результате самых разных предшествующих событий. Иными словами, мой мозг... мог иметь иную каузальную историю, нежели та, что действительно была у меня и составляла мою жизнь. Но если бы у меня была иная история, то, как следует из второй посылки, я обладал бы сейчас другими воспоминаниями и эти другие воспоминания фундировали бы иные убеждения и желания. Поскольку же мои убеждения и желания... скоррелированы с моим поведением (третья посылка), то, обладая другими убеждениями и желаниями, я вел бы себя иначе, чем веду себя в данный момент...» [Васильев, 2009: 218]. Итак, из аргумента следует, что, находясь в том же самом физическом состоянии, субъект мог бы вести себя различным образом. Эти различия обуславливаются только его квалиативными ментальными содержаниями, различными убеждениями и желаниями, и никакой анализ наличного физического состояния не способен дать точного прогноза действий субъекта. Ментальные состояния действительны, а значит, эпифеноменализм ложен.

Но опровержение эпифеноменализма, с моей точки зрения, не самое критичное из следствий аргумента каузальных траекторий. Гораздо более значимым кажется опровержение того, что является общим местом современных теорий сознания, – тезиса о супервертности (ТС) ментальных содержаний над состояниями мозга.

Возможны как минимум три формулировки тезиса о супервертности: (1) Ментальное супервертно над физическим в том случае, когда если нечто  $X$  имеет ментальное свойство  $M$ , то имеет физическое свойство  $P$ , такое, что  $X$  имеет  $P$ , и необходимо, что всякий объект, обладающий  $P$ , наделен  $M$ ; (2) Ментальное супервертно над физическим в том смысле, что объекты, совершенно одинаковые в своих физических свойствах, не могут отличаться относительно ментальных свойств; (3) Ментальное супервертно над физическим в том смысле, что миры, одинаковые во всех физических отношениях, одинаковы и во всех ментальных отношениях, т.е. миры, одинаковые физически, совершенно одинаковы в целом [Kim, 2006]. Сила этих утверждений различна. Наиболее сильным является утверждение (1), так как оно фиксирует супервертность даже одного изолированного свойства, а наиболее слабым – утверждение (3), так как в нем утверждается супервертность только миров в целом. Если условия, описанные в утверждении (1), выполняются, то будут выполняться и условия (2). А если выполняются условия по утверждению (2), автоматически бу-



дуг выполнены и условия по утверждению (3). Определение (3) фиксирует отношение глобальной супервертности, а определение (1)–(2) – локальной. Итак, как соотносятся три варианта тезиса с супервертностью с выводом аргумента каузальных траекторий?

Очевидно, реальная возможность существования двух идентичных людей с разными ментальными состояниями, допускаемая в качестве результата аргумента, противоречит варианту 1 ТС. Но сам автор сообщает, что его выводы не противоречат глобальной супервертности: «Последняя формулировка выражает так называемую глобальную супервертность, и мы увидим, что предложенные мной аргументы не отрицают ее. Вторая формулировка тоже не опровергается ими, если она допускает глобальную интерпретацию. Ошибочно лишь первое понимание» [Васильев, 2009: 243]. Что же получается? Перед нами неожиданная и удивительно красивая картина мира. Ментальные состояния отличаются от физических. При этом они не эпифеноменальны, а действительны. Они влияют на поведение. Но из-за этого влияния не разрушаются законы природы, в том числе фундаментальный закон сохранения энергии. Ментальное является посредником, условием влияния нелокальных физических факторов, а локально – причиной поведения. Ментальная каузальность защищена, тщетность усилий ученых найти корреляты ментальным состояниям объяснена, ПКЗ сохранен, и... в качестве бонуса получен ключ для решения двух философских проблем – «трудной проблемы сознания» и проблемы тождества личности. Из подобной картины следует, что сознательные состояния существуют потому, что необходимы для генерации человеческого поведения (без них поведение разумного человека было бы просто невозможно), и что тождество личности не связано только лишь с тождеством тела, а с тождеством каузальной истории, зафиксированной в ментальных состояниях. Что ж, если это построение помимо своих бесспорных эстетических качеств сможет выдержать критический анализ, это крупный успех, даже прорыв в философии. Итак, учитывая значимость выводов, вернемся к аргументу и попытаемся для начала проверить, действительно ли он опровергает тезис о локальной супервертности, сохраняя при этом глобальный вариант. Именно в этом случае он совместим с ПКЗ, вписывается в современную научную картину мира и согласуется с базовыми естественными установками.

### **Локальная или глобальная супервертность?**

Вполне вероятно, в исходном виде аргумент действительно направлен исключительно против локальной супервертности менталь-



ного над физическим. Но есть вероятность, что если аргумент верен, его вариацию можно использовать и против глобальной супервентности. Тогда рассуждения строились бы следующим образом. (1) У одинаковых событий могут быть разные причины. Таким образом, если текущее состояние мира представляет собой гигантский комплекс событий, эти же события могли произойти вследствие других причин. (2) Память людей обычно отражает события, которые происходили с ними в прошлом. Но если бы у них было другое прошлое, их память имела бы другое содержание. (В) Следовательно, возможно, что существовал бы мир, полностью идентичный нашему, но отличающийся от него содержанием ментальных состояний населявших его людей. Это следствие противоречит тезису о глобальной супервентности: тезису о том, что миры, одинаковые физически, совершенно одинаковы в целом. Выходит, что аргумент каузальных траекторий направлен и против глобальной супервентности?

Автор предполагает подобного рода замечания и заведомо не согласен с выводом. Скорее всего он критиковал бы первую посылку в нашем изложении. Хотя одно сингулярное событие (используя терминологию Васильева) может иметь множество предшествующих событий в качестве причины. Весь комплекс событий, составляющий состояния мира в момент  $t$ , мог бы иметь только одно предшествующее состояние, то, которое было в действительности, т.е. у текущего совокупного события – *События А* может быть только один antecedent – *Событие В* (ТСС – тезис о совокупном событии). Но каковы основания считать именно так?

У Васильева есть две тактики защиты этого тезиса. Одна основывается на анализе принципов здравого смысла. По мнению автора, эти положения не могут быть оспорены, так как отражают устройство наших познавательных способностей. Одним из базовых положений Васильев считает убеждение «в тождественности будущего прошлому». На нем построены другие: вера в существование причин у каждого события, убеждение в том, что «при точном повторении в будущем тех событий, которые имели место в прошлом, за ними последует то, что следовало за ними тогда, когда мы раньше ощущали их» [Васильев, 2009: 208], принцип каузальной замкнутости физического. По мнению Васильева, аксиома тождества прошлого и будущего обуславливает также веру в возможность существования только одной причины, только одного События А, предшествующего Событию В. Мне кажется, что следование здесь далеко не очевидное. Но в данном случае я не берусь оценивать умозаключения Васильева, так как метод философствования, который применяет автор, частично выходит за аналитический дискурс и представляет собой вариант аналитической феноменологии, довольно редко используемый современными философами. Впрочем, к выводам



автора проложен и другой путь, который проще проследить и проверить. Он представлен в самой новой книге В. Васильева «Сознание и вещи».

Философ сравнивает два описания события – отскок мяча от поверхности *simpliciter* и отскок мяча от поверхности на такое-то расстояние. Оба события могут быть следствием множества причин, например удара по мячу другим мячом, или теннисной ракеткой, или неустойчивого исходного положения мяча на плоскости. Но второе описание является более конкретным и в этом смысле сужает горизонт возможных причин (предшествующих состояний). Кажется, что конкретизация описания событий уменьшает возможные antecedentные события. Отскок мяча на 200 м мог иметь гораздо меньше причин, чем просто отскок мяча. А состояние мозга в момент  $t$  является еще более «конкретным событием», чем отскок мяча на 200 м, и потому может иметь еще меньшее количество причин. Но явно не одну. Только предельно конкретное событие имеет только одну причину. Таким событием и является Событие, или совокупное содержание всего чувственного опыта в момент  $t$ , считает Васильев.

Кажется, эти рассуждения не являются полностью безупречными. Они строятся на представлении о снижении количества возможных причин при конкретизации описания. Но разве они убеждают нас в том, что совокупное событие может иметь только одну причину? Возьмем сингулярное событие и создадим для него исчерпывающее конкретное описание. Это описание будет бесконечно длинным, пусть так. Но будет ли у этого конкретного события возможность в качестве antecedента иметь больше одной причины? Кажется, что да, по крайней мере в соответствии с посылкой (1) исходного аргумента. Вне зависимости от того, сколь точным, конкретным будет описание одного сингулярного события, к нему может быть больше одной каузальной траектории.

Теперь подумаем о Событии. Разве Событие, совокупное событие, текущее состояние мира, по определению не является совокупностью всех конкретных сингулярных событий? Если так, то его предшественниками могут быть состояния, состоящие из вариаций antecedentов всех конкретных сингулярных событий, и таких предшествующих состояний должно быть больше, чем одно. Я даже думаю: чем больше сингулярных конкретных событий входит в совокупность, тем больше вариаций может предшествовать. К примеру, комбинация шаров на бильярдном столе такова, что игрок может забить шар в лузу только с отскоком от правого или левого борта. Если шар уже попал в лузу и мы не видели, как он был забит, мы должны допустить, что он туда попал в результате одного из двух вариантов. Теперь представим, что футбольное поле или даже пустыня Сахара заставлены бильярдными столами с анало-



гичным расположением шаров и у каждого стола находится по одному игроку. Вот все шары оказались в лузах. Какое количество комбинаций могло предшествовать попаданию всех шаров в лузы? Очевидно, что это число больше, чем число комбинаций, предшествующих попаданию одного шара. Мне кажется, что таким же образом дело должно обстоять и с совокупным событием. Это значит, что если аргумент каузальных траекторий верен, он, вероятно, опровергает не только локальную супервентность, но и глобальную, т.е. все три варианта формулировки тезиса. В таком случае ментальное становится полностью автономным от физического, парящим свободно над физическим миром, а это нарушение ПКЗ, одного из принципов здравого смысла.

Теория, которая предполагает такие следствия, вряд ли может выстоять. Впрочем, не стоит торопиться. Даже если нам удалось опровергнуть второй путь защиты ТСС, достоверным мог быть первый. А для его анализа у нас недостаточно инструментов. Думаю, в таком случае оправданно вернуться к исходному аргументу и продумать его посылки. На мой взгляд, достаточно будет ограничиться первыми двумя. Ведь для получения интересующего нас вывода (о ложности тезиса о супервентности) достаточно только их. Мы сократим аргумент и приведем его в несколько адаптированный вид.

### **Разъяснение двусмысленности в аргументе. Преобразование аргумента**

(1) Реально возможно, чтобы мой мозг пришел к состоянию С в результате различных событий (А, В и т.д.).

(2) Реально возможно, чтобы память отражала события, происшедшие с субъектом в прошлом.

(В) Реально возможно, чтобы мой мозг был состоянии С, но моя память отражала другие события (А, В и т.п.).

В новой формулировке аргумента сделаны некоторые правки. Первая посылка конкретизирована в отношении возможных каузальных состояний мозга, ведь именно об этом в ней речь. Было введено уточнение: каждая из посылок представлена не как факт, а как реально возможный факт. Это кажется уместным дополнением, так как только в этом случае посылки станут верными. Мозг мог прийти к этому состоянию разными путями, но не приходил, а память реально может отражать прошлые события, но вовсе не всегда это делает. Вывод сокращенного аргумента опровергает ТС (1). Насколько я по-



нимаю, с такой формулировкой должен быть согласен автор аргумента каузальных траекторий<sup>2</sup>.

Однако в таком виде аргумента проявляется некоторая двусмысленность. Реально возможно, чтобы мозг пришел к состоянию С в результате разных событий, реально возможно, чтобы память отражала происшедшие с субъектом события. Но в двух посылках речь идет совсем о разных событиях: в одном случае о событиях, происходящих в нейронных сетях, а в другом – о биографических событиях. Память не отражает каузальных цепей нейронных событий, даже память нейрофизиолога-мнемониста, это реально невозможно. Тогда, значит, связь между посылками отсутствует? И из посылок вывода о ложности ТС вовсе не следует?

Конечно, дело обстоит не так. Двойственность из аргумента устранить возможно. Но тогда рассуждение должно иметь более сложный характер. Назовем это рассуждение АКТ<sup>3</sup>1:

(1) Реально возможно, чтобы мой мозг пришел к состоянию С в результате различных нейронных событий (А, В и т.д.).

(2) Другие нейронные события могли соответствовать другим биографическим событиям.

(3) Реально возможно, чтобы память отражала биографические события.

(В) Реально возможно, чтобы мой мозг был в состоянии С, но моя память отражала другие биографические события (А, В и т.п.).

Посылки представляются истинными. Следовательно, истинным может быть вывод. Правда, есть существенное условие. Для истинности вывода нужно не только, чтобы реально возможна была каждая из описанных в посылке ситуаций, но и их конъюнкция: все события должны быть возможны одновременно. Ведь реально возможно начать игру в шахматы с хода пешкой и с хода конем, но нельзя начать игру сразу с двух ходов – конем и пешкой<sup>4</sup>. Мне кажется, что совпадение ситуаций, описанных в посылках аргумента, все-таки невозможно. Невозможна ситуация, когда есть два двойника и их мозги находятся в идентичном состоянии, но пришли к этому состоянию в результате

<sup>2</sup> «Мой мозг... мог иметь иную каузальную историю, нежели та, что действительно была у меня и составляла мою жизнь. Но если бы у меня была иная история, то... я обладал бы сейчас другими воспоминаниями» [Васильев, 2009: 218].

<sup>3</sup> Попытаться выразить аргумент с помощью двух посылок тоже можно. Например, так: (1) Реально возможно, чтобы мое текущее физическое состояние имело другие причины. (2) Реально возможно, что память фиксирует историю моих физических состояний. (В) Реально возможно, чтобы у меня было текущее физическое состояние, но другие воспоминания. Мне кажется, что в таком виде он все еще сохраняет двойственность. Нужно объяснить. Коридор нейронных вариаций при одном биографическом событии. Незначимых для памяти физических состояний.

<sup>4</sup> Или: реально возможно, чтобы человек был выше среднего роста, но невозможно, чтобы все люди были выше среднего роста.



разных нейронных событий, а также разных биографических событий, и эти двойники имеют разные воспоминания.

## Возражение на АКТ1 – аргумент-двойник АКТ2

Доводом в пользу невозможности этой ситуации служит доказательство от противного. Далее я предложу аналогичный аргумент, аргумент-двойник, который хотя и имеет, по моему мнению, ключевые черты АКТ1, вероятно, приводит к ложному результату. Если аргумент-двойник не работает, но принципиально не отличается от АКТ1, то и АКТ1 не работает. Итак, проанализируем следующее рассуждение:

(1) Реально возможно, чтобы жесткий диск моего компьютера пришел в текущее состояние различными путями.

(2) Файловая система хранит информацию (компьютерная память) о последовательности изменения файлов на диске.

(В) Реально возможно, чтобы жесткий диск моего компьютера был в текущем состоянии, но файловая система хранила другую информацию о последовательности записи файлов на диск.

Это рассуждение похоже на АКТ1, и доводы, приводимые в пользу истинности посылок АКТ1, полностью применимы в новом аргументе. Диск компьютера – такая же физическая система, как мозг человека. Следовательно, текущему состоянию диска могли предшествовать различные состояния. Файловая система в качестве метаданных хранит информацию о времени создания, доступе и модификации каждого файла. Эта информация обычно представляется правильным образом и отражает индивидуальную каузальную траекторию жесткого диска. Отклонения, конечно, возможны, но мы говорим о норме. И это общее представление как любителей, так и профессиональных программистов. Тем не менее вряд ли кто-либо станет убеждать нас в отсутствии локальной супервентности данных файловой системы над физическим состоянием жесткого диска и допустит возможность ситуации, описанной в выводе АКТ2. Но почему? Почему вывод АКТ1 представляется правдоподобным в случае с памятью человека и неправдоподобным в АКТ2 в случае с памятью компьютера? Может, нам стоит поэкспериментировать с жесткими дисками и экспериментально проверить работу принципа супервентности в компьютерных системах?

Все-таки в случае с механическими системами в локальной супервентности информации над носителем вряд ли можно сомневаться. Здесь с нами согласится и Васильев. Об этом можно судить из его анализа чего-то подобного, робота-зомби с механической памятью.



«Такой робот мог прийти к своему нынешнему состоянию... разными путями. Но независимо от того, как он пришел к нему, он, будучи чисто механической конструкцией, должен в этом состоянии одинаково реагировать на свое окружение. Итак, для поведения нашего робота-зомби значима не его индивидуальная история, а исключительно его наличное состояние» [Васильев, 2009: 221]. Наличное состояние робота может отражать прошлые события. Но, по мнению Васильева, механическая память робота с таким же успехом могла отражать и отражает и другие прошлые события, потому что индивидуальная каузальная траектория в ней безразлична. На чем основана такая уверенность? Возьмем убеждение, что я родился в 1976 г., и запись компьютера о том, что первый файл на его диск записан в 1976 г. В чем принципиальная разница? Оба «воспоминания» содержат хронологическую метку, оба могут быть верными или ошибочными, оба в качестве основания имеют некоторый правдоподобный источник (часы реального времени и свидетельство о рождении) и оба влияют на поведение. Оба «воспоминания» могут как-то отражать другие события и оба могут исчезнуть или измениться, если оказать непосредственное воздействие на физический носитель, мозг или жесткий диск. Повреждения или значимые изменения в физическом носителе в обоих случаях приводят к изменениям или исчезновениям воспоминаний.

Может, различие в том, что механизм компьютерной памяти относительно хорошо изучен, а механизм работы человеческой памяти все еще отчасти остается загадочным для ученых? Но отсутствие понимания всех механизмов работы человеческой памяти не представляет собой в пользу наличия в ней нефизических свойств и отсутствия локальной супервентности. Может, в том, что память человека может содержать конкретные визуальные образы прошлого, а память компьютера – последовательность битов, которые могут быть проинтерпретированы самым различным образом? Например, запись о времени создания файла какой-то другой программой действительно может быть проинтерпретирована иначе, как математическая формула, данные об изображении, команда распечатать файл и т.п. С этим возражением можно работать. (Можно попытаться показать, что воспоминания человека могут также быть интерпретированы по-разному, как и любая сжатая информация. Тактики атаки и защиты подобных утверждений присутствуют, например, в дискуссии Сёрла и Деннета о китайской комнате.) Но более простой способ атаковать АКТ – заменить цифровой жесткий диск в аргументе-двойнике на аналоговый носитель, например аналоговую кассету или диафильм. Эти носители сохраняют информацию в максимально конкретном виде, визуальном, и альтернативная интерпретация данных затруднена. Мне кажется, что параллельный аргумент и его несостоятельность должны показать то, что сам по себе АКТ1 недостаточен для опровержения локальной су-





первентности. Для опровержения этого стороннику АКТ1 нужно показать, что аргументы неэквивалентны.

В рассуждениях Васильева есть указание на еще одно, возможно, ключевое отличие механической памяти от памяти человека. Видимо, именно оно должно делать АКТ1 правдоподобным, а его механическую вариацию АКТ2 – нет. В представлениях автора память человека – это приватные качественные состояния. Видимо, поэтому «индивидуальное прошлое не растворяется в нейтральных нейронных кодах, а продолжает существовать на уровне ментальных образов» [Васильев, 2009: 221]. Это утверждение нужно использовать как еще одну посылку в аргументе. Она фигурирует там проходящим образом как нечто само собой разумеющееся, но крайне важна и потому требует разъяснения и оценки.

Квалиа – термин, обозначающий качественный или феноменальный аспект ментальных процессов: мучительность боли, «красноту» красного, «горечь» обиды, «сладость» наслаждения. Его используют для того, чтобы провести различие между содержательным аспектом ментального процесса (контентом) и субъективным переживанием процесса, «каково это испытывать» процесс на себе. Когда Васильев говорит о том, что человеческая память состоит из приватных качественных состояний, он имеет в виду наличие этого качественного аспекта. Я вспоминаю прошлое лето: выходной день, дача, на траве стоят плетеный столик и кресло, на столе книги и чашка чая, почти ощущаю запах постриженной травы, слегка греет солнце, чувствую удовольствие с оттенком сожаления. Воспоминания как бы оживляют пережитые ощущения, мои личные, те, которые никто другой не переживал и о которых другие могут узнать только косвенным образом. Этот пример должен объяснить утверждение Васильева. Но если хорошо подумать, все ли содержимое памяти представлено таким же образом? Пусть некто знает иностранные языки, знает в смысле помнит тысячи слов, выражений и грамматические правила. Разве такое содержание памяти окрашено феноменально? Каково это – помнить три иностранных языка? И отличается ли это состояния от состояния помнить четыре или пять иностранных языков?

Когда я печатаю этот текст, я, очевидно, помню собственную биографию, но каково это – помнить свою биографию? Отличается ли это состояние от состояния «помнить биографию Стива Джобса или Бенжамина Франклина»? А было ли бы мое состояние другим, если бы вместо того чтобы сейчас помнить, что я родился Москве, я, например, помнил, что родился в Бейруте или Бостоне? Память об этих моментах, кажется, влияет на мое поведение. Но чтобы иметь кажимость влияния, они не должны быть как-то качественно окрашены.

Попробуем расставить все по местам. У нас есть АКТ1 и АКТ2. Единственным отличием между ними является допущение о том, что



память человека состоит из качественных частных состояний. Но на наших примерах очевидно, что хотя бы часть содержания памяти не имеет «качественного вкуса». Должны ли мы вопреки интуиции провести различие в каузальной действенности «качественно окрашенных» и «качественно пустых» воспоминаний? Или углубиться с интроспективные исследования, чтобы все-таки найти «вкус» всех без исключения записей в памяти? Мне кажется, что обе альтернативы бесперспективны и вернее считать, что память не имеет «качественных качеств». Тем более, что такую точку зрения могут занимать не только элиминативисты, но и такие квалифилы, как Д. Чалмерс.

### **Память не является качественным состоянием**

В «Сознающем уме» Чалмерс утверждает, что ментальное исчерпывается психологическим и феноменальным (качественным). Некоторые ментальные свойства сводятся только к психологическим, некоторые – только к феноменальным, а некоторые имеют двойственный характер. Боль, к примеру, имеет как психологические, так и феноменальные аспекты. Психологической составляющей является состояние, связанное обычно с повреждением организма и ведущее к особому поведению реакцией уклонения. Как мы видим, это описание функциональное – оно фиксирует входящие параметры и соответствующее этим параметрам поведение-реакцию. Феноменальным аспектом боли является ее мучительное переживание. Концептуально это несвязанные аспекты: теоретически организм мог бы быть поврежден и демонстрировать характерное для боли поведение, но при этом не испытывать болезненных переживаний и, наоборот, мог бы быть не поврежден и не демонстрировать боли в поведении, но интенсивно ее переживать. Фактически в мире и жизни организма они переплетены. А вот память Чалмерс относит целиком к психологическим процессам: «Согласно нашему анализу... обучение и память – функциональные свойства, характеризующиеся каузальными ролями...» [Чалмерс, 2013: 44]. Причина в том, что процесс запоминания и хранения в памяти можно легко представить бессознательным, лишенным феноменальных свойств. Она проходит своеобразный тест: если ментальное понятие *M* представимо без какого-либо связанного с ним феноменального качества, значит оно психологическое. В таком случае почему автор аргумента каузальных траекторий считает, что память находится в частных качественных состояниях?

Очевидно, Васильев, говоря о памяти, имеет в виду не процесс запоминания и хранения информации о прошлом, а его воспроизведе-



ние в текущем опыте субъекта. В текущем опыте часть воспоминаний действительно представляется феноменально окрашенной. Но тогда точнее будет сказать, что не память обладает качественными состояниями, а текущее состояние, объектом которого являются события из прошлого, из памяти. Содержание памяти может быть воспроизведено в сознании с феноменальными свойствами, но не имеет их в процессе хранения. Качественные аспекты проявляются в наличном состоянии субъекта, а не в его памяти.

Что если представить, будто эти ментальные содержания все-таки живут внутри нас в виде ментальных образов, постоянно текущих феноменальных состояний? Непонятно, как это представить, и непонятно, чьими бы эти состояния были. Если сейчас, когда я пишу эту статью, продолжает существовать феноменальное воспоминание моего воскресного летнего дня, то не совсем понятно, где и для кого оно продолжает существовать. Ведь онтология феноменальных состояний субъективна: если феноменальные состояния существуют, они существуют представленными для субъекта. Уверен, что в нейронных сетях закодированы воспоминания, но существование неосознаваемых никем феноменальных состояний мне представляется маловероятным. Тем более это трудно представить в связи с прерывностью сознательного опыта. Даже если прошлые феноменальные состояния продолжают существовать где-то на периферии моего сознания, то они должны улечиваться во время и после наркоза или даже сна. Но было бы странно предполагать, что только ментальные события после сна имеют каузальную действенность, а все, что происходило до сна, «растворилось в нейронных кодах». Чтобы защитить аргумент, не достаточно показать возможность «проигрывать» в настоящем воспоминания в «театре сознания». Нужны доводы в пользу непрерывного параллельного существования феноменальных состояний прошлого и непротиворечивый способ помыслить это. А этих доводов, судя по всему, мы пока не находим, и неосознанные сознательные состояния помыслить невозможно. Таким образом, различие в квалиа представляется мне нерелевантным и мы возвращаемся к исходной ситуации: если бы работал АКТ1, работал бы и АКТ2.

### **Решение парадокса сближающихся в ментальных состояниях двойников**

Впрочем, может быть, и тезис о супервентности ментального над физическим способен привести к трудностям подобного характера? Кажется, такое сомнение озвучивает Васильев, когда предлагает нам пред-



ставить двойников со сближающимися физическими состояниями. Вот в разных мирах существуют два идентичных человека, и их состояния постепенно сближаются. По мере этого сближения должны стираться и их ментальные различия. Если это напоминает амнезию, стирание тех аспектов прошлого, которые отличали их друг от друга, двойникам «придется забыть очень многое или даже почти все, если двойников очень много, а ведь это тоже реально возможно» [Васильев, 2009: 219]. Чем больше двойников с альтернативными траекториями, тем больше воспоминаний должно быть стерто. Чем больше двойников, тем меньше каждый из них может помнить. Человек, у которого бесконечное множество двойников, не должен помнить ничего. Значит, «сознанием в полной мере обладают только те люди, у которых нет двойников. Чем больше у вас двойников, тем больше вы и они должны быть похожи на зомби, хотя эти двойники никак с вами не связаны» [Васильев, 2009: 219]. Такое следствие действительно похоже на абсурд. Оно явно не согласуется со здравым смыслом и естественными установками. Неужели мы вынуждены принимать его, если отстаиваем тезис о супервентности и допускаем множественность каузальных траекторий?

Для того чтобы ответить на этот вопрос, думаю, сначала нужно представить ситуацию с двойниками с большими подробностями. Построим мысленный эксперимент с дополнительными деталями. Итак, есть два двойника, два полностью неотличимых по физическому строению человека, живущих в разных мирах. Пусть один из них, Дэн, родился в Бостоне, а другой, Фил, в Бейруте. И у них разные занятия: Дэн бизнесмен, а Фил музыкант. Один проводит все время в офисе, другой по вечерам играет на фортепиано в маленьком ресторанчике. В какой-то момент параллельных, но не пересекающихся жизней мозги двойников приходят в одинаковые состояния, одинаковые вплоть до каждого нейрона, атома, элементарной частицы. И в этот момент их спрашивают в каждом из миров: «Эй, ты кто? Чем ты здесь занимаешься?» По ТС их ответ должен быть идентичным. Предположим, они оба отвечают: «Чего тебе? Я музыкант, я играю на фортепиано». Но как такое помыслить? Как бизнесмен мог дать такой ответ? Неужели он забыл часы, проведенные в офисе? Он действительно должен был забыть многое, очень многое. Но очень многое он должен был и помнить. К примеру, он должен помнить мелодию и гармонии джазовых стандартов «Каравана» и «Мисти», неподражаемые виртуозные пассажи легендарного Арта Тейтума, тонко чувствовать свинговый ритм, синкопы, уметь импровизировать. Разве похоже все это на поведение и состояние зомби? Разве его нельзя назвать полноценным сознательным существом? Допустим, кроме Дэна и Фила были и другие двойники, и все они пришли к тому же состоянию, и все теперь говорят, что джаз – их призвание. Разве от этого сознание Дэна или Фила станет менее богатым? Вряд ли нам следует так думать.



Парадокс работает только в том случае, если допустить, что объемом возможных воспоминаний у двойников ограничен и что они только забывают, но не приобретают новые воспоминания. Только тогда это вело бы к сокращению возможного содержания памяти.

Действительно ли тезис о том, что двойники по мере схождения их физических состояний должны забывать отличающие аспекты прошлого, ведет в итоге к сокращению возможного содержания памяти? Мне кажется, это могло бы быть истинным только в случае конечного числа возможных воспоминаний. В самом деле, если от конечного числа возможных воспоминаний отнимать все большее и большее количество, будет оставаться меньше возможностей. Но если количество бесконечно, то от вычитания все большего и большего числа различий мощность исходного множества не уменьшится. Тогда, может быть, бесконечное число двойников будет представлять опасность для содержания памяти? Мне кажется, что и здесь не возникает проблемы. Ведь если из бесконечного множества натуральных чисел удалить бесконечное множество нечетных чисел, останется бесконечное множество четных и это множество по мощности будет таким же, как исходное множество.

Безусловно, допустимо оспаривать, что число возможных воспоминаний конечно. Я не вижу оснований так считать, но даже если это верно, допущенную нами ситуацию можно спасти от абсурда вторым способом. Дело в том, что двойники не только «забывают» отличающие их аспекты прошлого, но и приобретают новые воспоминания. Так, двойник-бизнесмен в нашей ситуации не только забыл часы, проведенные им в офисе, но, судя по всему, приобрел множество дополнительных воспоминаний, связанных с игрой на фортепиано. Содержание его памяти не уменьшилось, а только изменилось. И если физически состояние его мозга сближалось с состоянием мозга музыканта постепенно, стоит предположить, что менялось содержание его памяти. Постепенно он забывал офис и убеждался, что вместо офиса проводил время в ресторане и смотрел не на экран компьютера, а в ноты. Хотя все это вообразить очень сложно, имея представление о сложности устройства мозга, не менее сложно вообразить, чтобы два мозга оказались в абсолютно тождественном состоянии. Из приведенных выше рассуждений следует, что возможно непротиворечиво помыслить двойников со сближающимися ментальными траекториями, а это значит, что тезис о супервертности не ведет к противоречивым последствиям.

## Выводы

Тезис о супервертности ментального над физическим является традиционным в современной философии сознания. Аргумент АКТ1 направлен против этого тезиса, в частности против локальной супер-



вентности. В данной статье я защищаю ТС. В ходе исследования показано, что если аргумент подвергает сомнению локальную супервентность, он скорее всего подвергает сомнению и глобальную супервентность. Это своеобразное *reduction ad absurdum* аргумента. Однако данный вывод не окончателен. Более надежным является второй способ опровержения аргумента – доказательство от противного. В статье предложен новый аргумент АКТ2, сходный по своей структуре с АКТ1, но приводящий к очевидно ложным выводам. Если аргументы аналогичны, АКТ1 ложен. В статье рассмотрено возможное возражение – указание на то, что АКТ1 и АКТ2 различаются из-за отсутствия в АКТ2 квалиа. Однако в статье показано, что и в АКТ1 квалиа не имеют значения. В завершение представлено решение парадокса двойников с разными биографическими историями. Этот парадокс предложен Васильевым в качестве иллюстрации абсурдности следствий ТС. Однако более подробное рассмотрение этого мысленного эксперимента показывает, что тезис о локальной супервентности не ведет к абсурдным последствиям. Таким образом, представляется, что ТС защищен.

## Библиографический список

- Васильев, 2009 – *Васильев В.В.* Трудная проблема сознания. М., 2009.  
Васильев, 2014 – *Васильев В.В.* Сознание и вещи: очерк феноменалистической онтологии. М., 2014.  
Чалмерс, 2013 – *Чалмерс Д.* Сознательный ум. В поисках фундаментальной теории. М., 2013.  
Chalmers, 1993 – *Chalmers D.* Facing up the problem of consciousness // *Journal of Consciousness Studies*. 1995. Vol. 2, № 3. P. 200–219.  
Davidson, 2006 – *Davidson D.* Mental Events. The Essential Davidson. Oxford, 2006.  
Hare, 1984 – *Hare R.* Inagural Address: Supervenience // *Proceedings of the Aristotelian*. 1984. Vol. 58. P. 1–16.  
Kim, 1998 – *Kim J.* Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. Cambridge, 1998.  
Kim, 2006 – *Kim J.* Philosophy of Mind. Cambridge, MA, 2006.  
Kim, 2010 – *Kim J.* Essays in the Metaphysics of Mind. Oxford, 2010.  
Lewis, 1983 – *Lewis D.* Extrinsic Properties // *Philosophical Studies*. 1983. № 44. P. 197–200.  
Vasilyev, 2009 – *Vasilyev V.* “The Hard Problem of Consciousness” and Two Arguments for Interactionism // *Faith and Philosophy*. 2009. Vol. 26, № 5. P. 514–526.



## CAN WE SEE NATURAL KIND PROPERTIES?

**René Jagnow** –  
Associate professor,  
Department of  
Philosophy, University  
of Georgia. E-mail:  
rjagnow@uga.edu



Which properties can we visually experience? Some authors hold that we can experience only low-level properties such as color, illumination, shape, spatial location, and motion. Others believe that we can also experience high-level properties, such as *being a dog* or *being a pine tree*. On the basis of her method of phenomenal contrast, Susanna Siegel has recently defended the latter view. One of her central claims is that we can best account for certain phenomenal contrasts if we assume that we can visually experience natural kind properties. In this paper, I argue that certain kinds of low-level properties, namely shape-gestalt properties, can explain these phenomenal contrasts just as well as high-level properties. If successful, this is a modest, but nevertheless significant result. Even though it does not prove the falsity of Siegel's proposal, it nevertheless secures the existence of a plausible alternative.

**Key words:** Perceptual experience, sensory phenomenology, cognitive phenomenology, perceptual contents.

### 1. Introduction

Suppose you look at a dog. You see its color, shape, and location. But do you also see it as *being a dog*? Many philosophers hold that we can visually experience only low-level properties such as color, illumination, shape, spatial location, and motion. On this view – the *low-level view* – the property *being a dog* is not an aspect of your visual experience.<sup>1</sup> Others believe that we can also experience high-level properties, such as *being a dog* or *being a pine tree*. On this view – the *high-level view* – the property *being a dog* is an aspect of your visual experience.<sup>2</sup>

Susanna Siegel has recently presented a powerful defense of the high-level view based on her method of phenomenal contrast.<sup>3</sup> This method – an inference to the best explanation – takes a suitable phenomenal contrast between two experiences as explanandum and then argues that this contrast is best explained in terms of the

<sup>1</sup> It is difficult to draw the distinction between low-level properties and high-level properties in a precise way. For the argument in this paper, I will assume that color, shape, motion, location, and illumination properties belong into the former category and properties like *being a dog* into the latter. Versions of the low-level view can be found, for example, in Byrne (2009), Carruthers and Veillet (2011), Dretske (1995), Lyons (2005), Price (2006), Pautz (2009; 2010), Prinz (2006), Smith (2002), and Tye (1995). Lyons and Bayne call the low-level view the *conservative view*.

<sup>2</sup> Siegel calls the high-level view the *rich content view* and Lyons and Bayne call it the *liberal view*. Versions of the high-level view have been defended, for example, in Bayne (2009), Fish (2009), Johnston (2004; 2006), Masrour (2011), McDowell (1994), Nanay (2011), Searle (1983), Siegel (2006, 2007, 2010), and Siewert (1998).

<sup>3</sup> See: Siegel (2006, 2007, 2010). For a reply to criticisms of her argument, see: Siegel (2013).



high-level view. Using this method, Siegel has argued that viewers can visually experience a broad range of high-level properties, including natural kind properties (e.g., *being an apple*, *being a pine tree*), artificial kind properties (e.g., *being a cup*, *being a telephone*), semantic properties (word meanings), and emotional properties (e.g., *being angry*, *being concerned*).<sup>4</sup> A number of authors have attempted to reject Siegel's argument by presenting alternative explanations of the relevant phenomenal contrasts in terms of low-level properties.<sup>5</sup> Yet, since Siegel's argument is an inference to the best explanation, her critics can succeed in this only if they also show that their alternatives explain the relevant phenomenal contrasts at least as well as Siegel's high-level view.<sup>6</sup> This, in my view, is one of the most significant challenges posed by Siegel's argument.

In this paper, I will take up this challenge for certain phenomenal contrasts, namely those that, according to Siegel, are best explained in terms of the viewer's visual experience of natural kind properties.<sup>7</sup> I will first develop an alternative explanation of these phenomenal contrasts in terms of certain low-level properties, namely shape-gestalt properties. Siegel discusses and rejects this alternative (Siegel, 2006: 499–500; 2010: 110–113). But, as I will show, her arguments against this alternative are not conclusive. I will then argue that shape-gestalt properties can explain the phenomenal contrasts just as well as natural kind properties. If successful, this is a modest, but nevertheless significant result. Even though it does not prove the falsity of the high-level view with regard to natural kind properties, it nevertheless secures the existence of a plausible alternative.

The paper is organized as follows. In section 2, I clarify the low-level and the high-level view. In section 3, I explicate Siegel's method of phenomenal contrast and show that her arguments against explanations in terms of shape-gestalt properties are not conclusive. In section 4, I develop the shape-gestalt proposal in greater detail. The purpose of this section is to clarify the notion of a shape-gestalt property and to show that the shape-gestalt proposal is plausible from an empirical point of view. In section 5, I argue that shape-gestalt properties can explain the relevant phenomenal contrasts just as well as natural kind properties.

<sup>4</sup> Nanay (2011; 2012) has argued that the rich content view is false for natural and artificial kind properties, but true for dispositional properties such as *being climbable*, *being edible*, or *being Q-able* in general. I will not discuss his view in this paper.

<sup>5</sup> See, for example, Price (2009) and Prinz (2013). Related criticisms can be found in Logue (2013), Brogaard (2013), Pautz (2009), and Nanay (2011).

<sup>6</sup> Ideally, they would have to show that their explanations are better than Siegel's. Siegel writes, for example: "An abductive inference cannot be rebutted merely by the existence of alternative explanations. To rebut an abductive argument, the alternatives have to be put forward as being better than the explanation they are alternatives to" (Siegel, 2013: 854). However, in this paper, I will be concerned only with the more modest task. The reason for this will emerge in section 5 of this paper.

<sup>7</sup> Note that that some low-level properties, such a color properties, might be construed as natural kind properties. In this paper, I will therefore define the high-level view in terms of natural kind properties that are not plausibly also low-level properties.





## 2. The low-level view and the high-level view

We can formulate the two opposing views – the high-level view and the low-level view – more precisely by appeal to three distinctions. The first distinction concerns the nature of perceptual mental states. Many philosophers, including Siegel, argue that perceptual states have representational contents, that is, they represent the world as being a certain way. On this view – the *content view* – to see the dog's color is to represent the dog as *being brown*.<sup>8</sup> Other philosophers reject the idea that perceptual states have contents. On this view – the *relational view* – to see a property is to stand in a specific relation to that property, for example, the relation of *being sensitive to* or *presenting* the property.<sup>9</sup> The distinction between the low-level view and the high-level view can be formulated in both frameworks. But, for the purpose of this paper, I will follow Siegel and adopt the content view.

The second distinction concerns the notion of phenomenal character. I use the expression *phenomenal character* in the sense originally introduced by Nagel: to say that a mental state has phenomenal character is to say that there is something it is like for the subject to be in that state.<sup>10</sup> Various authors believe that it is helpful to distinguish between sensory and cognitive phenomenal character. Although there is no agreement on a set of criteria that would allow us to decide whether a specific phenomenal character is sensory or cognitive, we can get a rough idea by looking at examples. When you see the dog's color, your experience has a brownish sensory phenomenal character. In contrast, when you merely think that the dog is brown without seeing or imagining its color, your experience is missing this brownish sensory phenomenal character. Yet, this does not mean that your thought lacks any phenomenal character. Rather, as some have argued, occurrent thoughts may have cognitive phenomenal character.<sup>11</sup> According to them, there is something it is like for a thinker to engage in the activity of thinking, or there is something it is like for a thinker to entertain a certain proposition.

The third distinction concerns the difference between perceptual and non-perceptual states. The latter include feelings, beliefs, desires, judgments, and so on. There are different ways in which this distinction can be drawn. Some have argued that perceptual states differ from other types of

<sup>8</sup> The content view has been defended, for example, in Peacocke (1992), Crane (2006), Siegel (2010), Pautz (2009, 2010), and Schellenberg (2011).

<sup>9</sup> The relational view has been defended, for example, in Campbell (2002), Travis (2004), Martin (2004), and Brewer (2006).

<sup>10</sup> Nagel (1974).

<sup>11</sup> See, for example, Strawson (1994, 2011), Siewert (1998, 2011), Horgan and Tienson (2002), Tienson (2011), and Pitt (2004, 2011). For discussion of cognitive phenomenal character, see Siegel (2006, 492–497; 2010, 91–94, 102–108).



mental states in virtue of their specific phenomenology. Others have argued that perceptual states differ from non-perceptual states in terms of their function. For the purpose of this paper, I will follow Lyons (2005) who defines perceptual states in a non-phenomenological manner as output states of dedicated perceptual systems. For example, a visual perceptual state is an output state of the visual system. According to Lyons, such output states are typically complex states that contain sensations and experiential states. In addition, they might also contain non-experiential states, that is, high-level categorizations, such as representations of objects as having certain properties (Lyons 2005, 242).<sup>12</sup>

With these three distinctions in place, we can now formulate the low- and the high-level view in a more precise way. Let us say that a *visual experience* is that component of a visual perceptual state that is determined by its sensory phenomenal character. Let us also say that the *sensory phenomenal content* of a visual experience is that aspect of its representational content that supervenes on its phenomenal character. We can then formulate the low- and the high-level view with regard to visually recognizable natural kind properties as follows:

- Low-level view*    *Only low-level properties (e.g., color, shape, and location properties) enter into the sensory phenomenal contents of visual experiences.*
- High-level view*    Sometimes, at least some natural kind properties (e.g., being a dog, being a pine tree) enter into the sensory phenomenal contents of visual experiences.<sup>13</sup>

Formulated in this way, the issue at hand is about the scope of the sensory phenomenal contents of visual experiences.

Before turning to Siegel's argument, I want to make the following comment. It might turn out that visual representations are distributed representations that bridge areas of the visual system with higher cortical areas. Since this would plausibly undermine the distinction between visual experiences and other mental states, it would then no longer make sense to ask

<sup>12</sup> This has an important consequence. It is possible that visual perceptual states have both sensory and cognitive phenomenal character. Indeed, this kind of view has been defended in Reiland (2014).

<sup>13</sup> Without the restriction to natural kind properties, we would have to define the high-level view as follows: Sometimes, at least some high-level properties enter into the sensory phenomenal content of visual experiences. Siegel is sometimes interpreted as holding a different view. For example, Bayne and Montague suggest that Siegel claims that high-level properties are represented by non-sensory, cognitive phenomenal aspects of perceptual states. See, for example, Bayne and Montague (2011: 22). But this does not seem to be a correct interpretation of her argument. When explaining her version of the high-level view, which she calls "Thesis K," Siegel writes, for example: "Thesis K amounts to this: in whatever sense the representation of color and shape properties can have an associated sensory phenomenology, the representation of K-properties [high-level properties] can too (Siegel, 2006: 485).



which properties can be represented by the former. For the argument in this paper, I will assume that this is not the case.

### 3. Siegel's argument for the high-level view

How can we decide whether the low- or the high-level view is correct? One might think that the best way to do so is through introspection. If the property *being a dog* were part of the sensory phenomenal content of your experience, there would be a difference between what it would be like for you to undergo this experience and what it would be like for you to undergo an experience of a dog replica, that is, an experience of something that has the same shape and color as a dog, but is not represented in the experience as such. Moreover, if the two experiences differed in phenomenal character because one, but not the other, represented the property *being a dog*, introspection should be able to tell you that this was the case. Yet, pure introspection, that is, introspection unaided by argument, has proven to be a very unreliable guide to the contents of perceptual experiences.<sup>14</sup> Even in the case of properties such as shape and color, introspection often does not yield decisive results. Consider, for example, the famous question of whether a coin seen from an angle looks round or elliptical. As the ongoing debate among philosophers shows introspection is not able to yield a satisfactory consensus on this question.<sup>15</sup>

In order to circumvent the problem resulting from the unreliable character of introspection, Siegel construes the method of phenomenal contrast as an inference to the best explanation. Such an inference proceeds in two steps. The first step presents two experiences that are such that subjects who undergo them typically agree that they differ in phenomenal character. The purpose of this step is to provide a noncontroversial, minimal phenomenal contrast that can serve as a suitable explanandum for the second step of the method. The second step then shows that the difference in phenomenal character is best explained by assuming that some natural kind property enters into the sensory phenomenal content of one experience, but not into the sensory phenomenal content of the other.

Siegel provides a number of phenomenal contrasts. In this paper, I will focus on her well-known pine tree example because it concerns natural kind properties.<sup>16</sup> Suppose you are given the task of cutting down all the

<sup>14</sup> For critical assessments of pure introspection, see, for example, Prinz (2006), Kriegel (2007), Schwitzgebel (2008), and Bayne (2009).

<sup>15</sup> Some philosophers have argued that the coin looks elliptical, some have argued that it looks round, and others again have argued that it looks both round and elliptical. For two classical versions of the first two views, see Russell (1912) and Austin (1962). An important defense of the third view can be found in Noë (2004).

<sup>16</sup> For this example, see: Siegel (2006: 491; 2010: 100).



pine trees in a forest. At first, you are unable to distinguish the pine trees from other similar looking trees. But, after repeated observations, you acquire the disposition to recognize a pine tree when you see one. Now consider two experiences, *O1* and *O2*. *O1* is your overall experience when you look at the pine tree before you acquire the recognitional disposition and *O2* is your overall experience after you acquire the recognitional disposition. We will stipulate that you look at the tree from the same point of view and under the same conditions of illumination. In this case, there is a wide consensus that *O1* and *O2* differ in their overall phenomenal character.

Let me emphasize two things about this phenomenal contrast. First, *O1* and *O2* are the overall experiences you have on the two occasions. They are temporally extended slices of your stream of consciousness. *O1* and *O2* will include as parts two visual experiences of the tree, call them *V1* (your visual experience of the tree before you acquire the recognitional disposition) and *V2* (your visual experience of the tree after you acquire the recognitional disposition). But *O1* and *O2* will also include other mental states. For example, *O2* might include a recognitional state, perhaps the belief that the object in front of you is a pine tree.

Second, according to Siegel, we can rely on introspection to tell us that *O1* and *O2* differ in their overall phenomenal character, but we cannot rely on introspection to say anything more specific about the nature of this difference. Introspection does not tell us whether this is a contrast in sensory or cognitive phenomenology. And it also does not tell us whether the contrast concerns the spatial phenomenology of the two experiences.<sup>17</sup>

*O1* and *O2* provide a suitable phenomenal contrast for the first step of the method. It is antecedently plausible that the two visual experiences contained in them – *V1* and *V2* – represent identical low-level properties. At least initially, there is little reason to think that they represent the pine tree as instantiating different color and shape properties.<sup>18</sup> Moreover, the two experiences differ in such a way that one, but not the other, allows you to recognize that the perceived object instantiates a natural kind property, namely the property *being a pine tree*. Thus, unless there is a better explanation, it is plausible that the phenomenal difference between *O1* and *O2* is due to the fact that the property *being a pine tree* enters into the sensory phenomenal content of *V2*, but not into the sensory phenomenal content of *V1*.

The task of the second step of the method of phenomenal contrast is to show that an explanation of the phenomenal contrast between *O1* and *O2* in terms of natural kind properties is better than its low-level alternatives. In order to do so, Siegel rejects three alternative low-level explanations. The third alternative appeals to shape-gestalt properties and is therefore central to this paper. Nonetheless, it will be helpful to at least briefly characterize the other two alternatives.

<sup>17</sup> See Siegel (2007: 130; 2010: 80–83).

<sup>18</sup> Yet, as we will see in the next section of this paper, this assumption is most likely false.



## CAN WE SEE NATURAL KIND PROPERTIES?

The first low-level alternative is to insist that *V1* and *V2* have identical sensory phenomenal character and try to explain the phenomenal difference between *O1* and *O2* in terms of some kind of non-sensory phenomenal character of other mental states contained in *O1* and *O2*. There are a number of options here. One option is that the phenomenal difference between *O1* and *O2* is a result of the fact that the latter, but not the former, contains a recognitional state with its own distinct cognitive phenomenology. Another option is that the difference is a difference in background phenomenology, say, for example, a difference in mood.<sup>19</sup> The second low-level alternative is to admit that *V1* and *V2* differ in phenomenal character, but maintain that this is not a difference in their representational contents. Here there are many possible options. One option is that *V2*, but not *V1*, involves a nonrepresentational feeling of familiarity. For the purpose of this paper, I will grant that Siegel's arguments against these alternatives are successful.

Let us now turn to the third alternative. The proponent of the low-level view could say that *O1* and *O2* differ in phenomenal character because *V1* and *V2* represent different low-level properties. Siegel thinks that an explanation in terms of shape-gestalt properties is the most plausible way of cashing out this alternative: "When you learn to recognize pine trees by sight, your experience comes to represent a complex of shapes – leaf shape, trunk shape, branch shape, and overall pine-tree shape. This complex is an overall pine-tree gestalt" (Siegel, 2010: 111). According to this proposal, *V1* and *V2* differ in phenomenal character because the latter, but not the former, represents a complex shape property, namely a pine-tree shape-gestalt. Since such a property is plausibly a low-level property, there is no need to appeal to natural kind properties in order to explain the phenomenal contrast.

Siegel puts forward two arguments against this low-level proposal.<sup>20</sup> The first argument is contained in the following passage:

[S] hape-gestalts that are abstract enough to remain invariant across pine trees will be invariant across other objects as well. For instance, a typical handgun, a drill, and a hair drier share the same shape-gestalt. The more abstract the shape-gestalt is, the less reason there is to think that experience fails to represent it prior to one's gaining a recognitional disposition. And if it is represented in experience prior to gaining the recognitional disposition, then it cannot explain the phenomenal contrast (Siegel, 2010: 111–112).

<sup>19</sup> As I pointed out above, it is possible that perceptual states contain cognitive states. In this case, Siegel would also have to show that the difference between *O1* and *O2* could not be explained in terms of the cognitive phenomenal character of a possible recognitional state contained in *V2*, but not *V1*.

<sup>20</sup> Note, however, that Siegel does not consider these arguments to be decisive [Siegel, 2010: 111].



This argument has the following two premises: (i) shape-gestalt properties that are abstract enough to remain invariant across pine trees will be invariant across other objects as well, and (ii) shape-gestalt properties that are abstract enough to remain invariant across pine trees are represented in experience prior to one's gaining a recognitional disposition. I have worries about both of these premises. As we will see later on, shape-gestalt properties can be very specific. This raises problems for the first premise. But I also think that the second premise is possibly false. It is plausible that one can acquire the disposition to visually recognize pine trees only together with the disposition to detect the corresponding shape-gestalt property. So, even if a given shape-gestalt property were instantiated in different objects, say, in a handgun, a drill, and a hair drier, this would not imply that viewers could detect this property prior to being able to recognize these objects. Siegel would have to exclude this option.<sup>21</sup>

Siegel's second argument is that the strategy of invoking representations of shape-gestalts is not generally available to the proponent of the low-level view. In order to make this case, she considers the phenomenal contrast between two experiences of a face that expresses doubt. One experience takes place before and the other after one learns that this particular facial expression expresses doubt. The proponent of the high-level view would say that the two experiences differ in phenomenal character because they represent different emotional properties. In order to respond to this, the proponent of the gestalt-proposal would have to show that the two experiences represent different gestalt-properties. But Siegel rejects this. She holds that "in this sort of case, it seems implausible to suppose that there must be a change in which color and shape properties are represented" (Siegel, 2010: 112).

Yet, as it stands, this example is not a conclusive response to the shape-gestalt proposal for at least three reasons. First, once one considers actual, rather than imagined, experiences, one might *discover* that there is indeed a phenomenal difference between the two experiences of the face, a difference that is due to the fact that they represent different shape-gestalts. Second, one might argue that if there is no difference with respect to low-level properties, there might simply not be a difference in sensory phenomenal character at all. It is possible that the two overall experiences differ in cognitive or emotional phenomenal character. Finally, even if emotional properties enter into the sensory phenomenal contents of visual experiences, it does not follow that natural kind properties do so too. To show that her argument is conclusive, Siegel would have to exclude these three possibilities.

<sup>21</sup> For other criticisms of this part of Siegel's argument; see Price (2009), Nanay (2011), Logue (2013), and Prinz (2013).



#### 4. The shape-gestalt proposal

As we have seen, there are no conclusive reasons for dismissing the shape-gestalt proposal. I will now develop the shape-gestalt proposal in greater detail. The main goal here is to get a better understanding of the notion of a shape-gestalt and to show that the shape-gestalt proposal is plausible from an empirical point of view. The notion of a shape-gestalt derives in part from observations of gestalt-switches afforded by ambiguous figures. Considering certain types of ambiguous figures therefore provides a suitable point of departure.

Ambiguous, or reversible, figures can be categorized in terms of the types of features that reverse. The duck/rabbit figure (Fig.1) can be seen as a duck and as a rabbit and is thus *content reversible*.<sup>22</sup> The Necker-cube (Fig. 2) can be seen from two different perspectives and is thus *perspective reversible*. Rubin's vase (Fig. 3) is *figure/ground reversible*. I will consider the *chef/dog figure* (Fig. 4), which, like the duck/rabbit figure, is exclusively content reversible, but, unlike the duck/rabbit figure, does not contain any parts that can be recognized independently of the entire figure. Note that the chef/dog figure is not one of a kind. Fig. 5 shows a similar figure, the so-called *whale/kangaroo*.

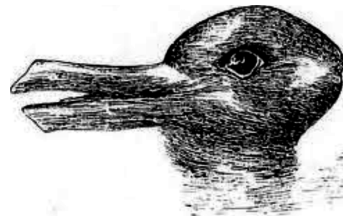


Fig. 1. Duck/rabbit figure

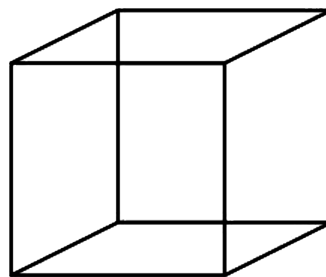


Fig. 2. Necker-cube

<sup>22</sup> A viewer who is familiar with both ducks and rabbits will experience the figure either as a duck or as a rabbit. Moreover, if the viewer knows that the figure can be seen in two different ways, her experience will reverse, more or less automatically, from one to the other. There is quite a bit of empirical evidence that perceptual reversal usually requires knowledge that the figure is reversible. See, for example, Rock and Mitchener (1992), Rock, Gopnik, Hall (1994), and Gopnik and Rosati (2001).

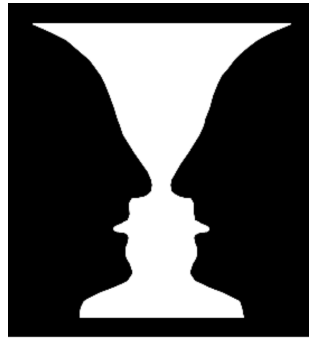


Fig. 3. Rubin's Vase

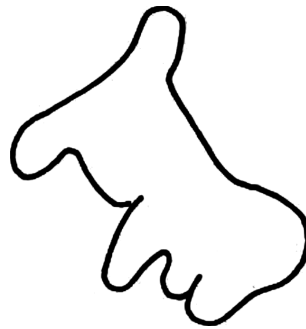


Fig. 4. Chef/dog figure

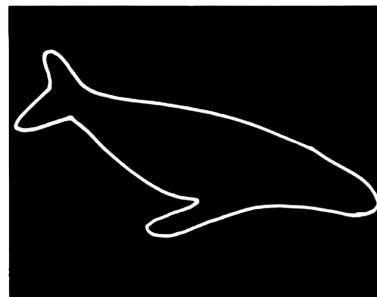


Fig. 5. Whale/kangaroo figure

The visual experiences elicited by the chef/dog figure can serve as explanandum for the first step of the method of phenomenal contrast. There is a clear phenomenal difference between your visual experience of the chef/dog figure as a chef (call this  $V_c$ ) and your experience of the figure as a dog (call this  $V_d$ ). The proponent of the high-level view could argue that  $V_c$  and  $V_d$  represent the same low level-properties, the same colors and shapes, and account for the phenomenal difference by saying that  $V_d$  represents the figure as instantiating the natural kind property *being a dog* and





Vc represents the figure as instantiating the natural kind property *being a chef*.<sup>23</sup>

The proponent of the low-level view could respond by saying that Vc and Vd differ in phenomenal character because they represent different shape-gestalt properties. But before we can understand how this might work, we need to look more closely at two features of the gestalt-switch. First, Vc and Vd organize the figure into two different part-whole structures. This becomes particularly obvious when we consider the relationship between parts and subparts. Consider, for example, the part that comes to be seen as the chef's hat. Whereas Vc represents the hat as one single part of the figure with a complex shape, Vd represents it as consisting of three subparts: leg, tail, and part of the dog's trunk. Similarly, when your experience switches from Vc to Vd, the same protrusion is first seen as a nose, that is, as a subpart of the chef's face, and then as leg, that is, as a subpart of the dog's trunk.

Second, the gestalt-switch is holistic. When you see the figure as a dog, you see all of its parts as parts of a dog (as dog-head, dog-leg, dog-tail, and so on) and when you see the figure as a chef you see all of its parts as parts of a chef (as chef-head, chef-nose, chef-chin, and so on). The important point for us is that you cannot see the parts of the figure as parts of a dog or a chef without seeing the figure as a whole as a dog or a chef. Consider, for example, the protrusions that come to be seen as the dog's legs. If you look at them in isolation, you cannot recognize them as being leg-shaped. This holds for the entire figure – all of its parts are strongly underdetermined.

The proponent of the low-level view has to construe the respective shape-gestalt properties in such a way that they can account for these two features. The obvious way to do so is to construe shape-gestalt properties as complex properties that specify both abstract part-whole structures and abstract outline shapes of the whole and its parts. If we conceive of shape-gestalt properties in this way, it is no longer plausible to say that “a typical handgun, a drill, and a hair drier share the same shape-gestalt,” as Siegel suggested [Siegel, 2010: 111]. Perhaps, there is a sense in which these objects all share an abstract outline shape. But they plausibly differ with regard to both their part-whole structures and the outline shapes of their parts.

Is the shape-gestalt proposal plausible from an empirical point of view? The figure is just an outline. Thus, in order to be able to see it as instantiating one of the two shape-gestalt properties, your visual system has to organize it into the respective part-whole structure in a segmentation process that uses stored information about the object's structure.<sup>24</sup> In other

<sup>23</sup> Strictly speaking, the natural kind property would be *being a human being*. But for simplicity, I will talk about the property *being a chef*. Some high-level theorists have used ambiguous figures as evidence for their view. See, for example, Searle (1983).

<sup>24</sup> In the literature on perceptual organization, this process is often called *perceptual parsing*, that is, the segmentation of an object into parts.



words, the proposal requires both that information about an object's likely gestalt is stored in memory and that this information can travel down from higher-level memory processes to the lower-level segmentation processes that organize the perceptual image. This is indeed consistent with recent research about perceptual organization. Many vision scientists believe that the visual system uses memories of object structure in the process of perceptual organization.<sup>25</sup> The existence of back projection fibers from higher levels of the visual pathway to lower levels that would make this possible is well established.<sup>26</sup> And it has also been shown that information travels back from areas in the inferior temporal cortex – an area usually associated with memory representations – to prestriate areas – that is, to areas where mental images are formed.<sup>27</sup>

Can the shape-gestalt proposal plausibly be extended to examples like Siegel's pine tree? One worry derives from the following asymmetry. In the case of the chef/dog figure, the viewer switches back and forth between the two contrasting visual experiences with the recognitional dispositions already in place. In the case of the pine tree, the viewer's experience changes over time as she acquires the recognitional disposition. In my view it is plausible to assume that viewers acquire the information that allows them to detect the relevant shape-gestalt properties of objects in a piecemeal fashion through repeated observation. After all, trees have relatively many parts that differ from those of other, similar looking trees in subtle details. Once this information has been acquired and stored in memory as a single representation, it can then travel downwards to low-level segregation processes, thus making it possible to detect the relevant shape-gestalt property.<sup>28</sup>

A second worry derives from the fact that experiences of ambiguous figures are experiences of pictures of objects rather than experiences of objects in the viewer's environment. I want to concede that when one sees the chef/dog figure, say, as a dog, one usually undergoes a visual experience that is specific to pictures. Such an experience might involve one's awareness that one is looking at a picture of a dog, rather than at a real dog, and it might also involve one's awareness of various features of the picture surface as surface properties. Nonetheless, many philosophers have argued that an experience as of an object in a picture resembles, in important respects, a face-to-face experience of the object represented in

<sup>25</sup> Helpful overviews of the relevant literature can be found in Kimchi, et al. (2003). The papers by Palmer, Peterson, Kimchi, and Kellman are particularly relevant here.

<sup>26</sup> See, for example, Lamme and Roelfsema (2000).

<sup>27</sup> See, for example, Kosslyn (1994).

<sup>28</sup> One can support this proposal with examples from perceptual training. One example can be found in Gauthier and Tarr (1997). For their experiments, Gauthier and Tarr trained subjects to become experts at recognizing novel objects called *greebles*. Greebles are unfamiliar complex visual objects that have typical distinctive parts that allow subjects to distinguish between different individual greebles, and to categorize them according to gender and family. The important point for us here is that the training process was very arduous.



## CAN WE SEE NATURAL KIND PROPERTIES?

the picture.<sup>29</sup> If this is correct, it is also plausible to assume that pictorial experiences employ, at least in part, the very same perceptual mechanisms as ordinary face-to-face experiences.

A third worry derives from the fact that some authors have urged that we distinguish seeing from seeing-as.<sup>30</sup> I think here we can admit that a visual experience of a gestalt-switch differs phenomenally from an ordinary experience of, say, a pine tree. In contrast to the latter, the former involves a change from one visual interpretation to another. However, this is entirely consistent with the claim that seeing a pine tree is similar to seeing, say, the chef/dog figure as a dog, in that both involve the detection of a specific shape-gestalt property. In my view, the best way to realize this is by appeal to three-dimensional versions of ambiguous figures. Consider, for example, a three-dimensional version of the duck/rabbit. If you look at such a figure, your experience will switch between two different face-to-face experiences of an ordinary three-dimensional object in your environment. Here, I think, it is very plausible to assume that each of these two experiences is an ordinary recognitional experience with the same structure as a recognitional experience of a pine tree.

Does the shape-gestalt proposal present a distinct alternative to the high-level view? It seems plausible that we sometimes sort objects into types according to their appearances. The proponent of the high-level view could therefore suggest that the relevant natural kind properties are ‘qualitatively pure’ in the sense that they are defined in terms of a distinct look, say, for example, a pine tree-like appearance. She could further suggest that that we define such appearances in terms of shape-gestalt properties. In this case, there would be no difference between a visual experience that attributes a qualitatively pure natural kind property (e.g., *being a pine tree*) to an object and a visual experience that attributes the corresponding shape-gestalt property (e.g., *having a pine tree shape-gestalt*) to that object, that is, there would be no difference between the high-level view and the shape-gestalt proposal. But, as Brogaard has pointed out, purely qualitative kinds are defined by their low-level properties and thus “are not really high-level properties and hence not really natural kind properties of the sort we are interested in here” [Brogaard, 2013: 40].

Is the shape-gestalt proposal plausible from a phenomenological point of view? The proponent of the high-level view might argue that the shape-gestalt proposal fails to account for a number of obvious phenomenal contrasts.<sup>31</sup> Consider the following scenario. As you are walking through a forest, you see a snake in front of you. Let us call the overall expe-

<sup>29</sup> For different views along these lines, see Gombrich (1960), Schier (1986), and Lopes (2006).

<sup>30</sup> In the context of his version of the low-level view, Prinz has insisted on this distinction. See: Prinz (2006: 436–437).

<sup>31</sup> Thanks to Igor Gasparov for pressing this point and for the following example.



rience at that time  $O_s$  and the visual experience of the snake contained in it  $V_s$ . After the initial shock, you notice that there is no snake in front of you, but rather a longish inanimate object with a snake-gestalt. Let us call your overall experience at that time  $O_i$  and the visual experience contained in it  $V_i$ . In this scenario, it is plausible that the two overall experiences differ in phenomenal character. Given this, the proponent of the high-level view could adapt some of Siegel's arguments in order to show that it is not possible to explain the phenomenal difference between  $O_s$  and  $O_i$  either in terms of the non-sensory phenomenal character of the non-visual states contained in them or in terms of some kind of background phenomenology. Since  $V_s$  and  $V_i$  attribute the same shape-gestalt property to the object, it then seems to follow that the shape-gestalt proposal is unable to account for the phenomenal difference between  $O_s$  and  $O_i$ .

Yet, the shape-gestalt proposal can account for this phenomenal difference in the following way. According to the framework outlined in the second section of this paper, the two visual experiences  $V_s$  and  $V_i$  are components of visual perceptual states that might also contain non-experiential categorizations. The perceptual state contained in  $O_s$  could categorize the object as a snake and the perceptual state contained in  $O_i$  could categorize it as an inanimate object. Moreover, it is possible that these categorizations differ in cognitive phenomenal character. In this case,  $O_s$  and  $O_i$  would differ in phenomenal character because the visual perceptual states contained in them would differ in cognitive phenomenology.<sup>32</sup>

## 5. Explaining the phenomenal contrasts in terms of shape-gestalts

In this section, I argue that shape-gestalt properties can explain the relevant phenomenal contrasts just as well as natural kind properties. But before doing this, I would like to clarify the task at hand in a bit more detail.

Inferences to the best explanation require criteria that allow us to assess how well different possible explanations account for the phenomena in question. Scientific explanations are often compared with regard to their predictive power, their ability to unify a large number of different phenomena, and their simplicity. But these criteria are not applicable to the present situation. The high-level view and the low-level view are not formulated as scientific theories that make empirically testable predictions. More-

<sup>32</sup> Scenarios like this one could threaten the shape-gestalt proposal only if there was a clear difference in sensory phenomenal character between the two visual experiences contained in them. In section five of this paper, I will discuss a similar scenario, the salamander-scenario, in order to show that there is no obvious difference in sensory phenomenal character between the two visual experiences involved in this scenario and I will give reasons for thinking that this holds in general.



## CAN WE SEE NATURAL KIND PROPERTIES?

over, both views claim to be able to account for the same range of phenomena, and it is hardly plausible to discount one of them because of its lack of simplicity. I will therefore not discuss how well the shape-gestalt proposal satisfies a given set of criteria. Rather, I will compare the explanatory power of the shape-gestalt proposal with that of the high-level view in a more direct way.

For this purpose, I will proceed in two steps. I first argue that we can provide explanations in terms of shape-gestalt properties for all those phenomenal contrasts that can plausibly be explained by appeal to natural kind properties. The purpose of this step is to secure the existence of appropriate low-level explanations. I then argue that shape-gestalt properties can explain these phenomenal contrasts just as well as natural kind properties. The purpose of this second step is to show that explanations in terms of shape-gestalt properties are phenomenally adequate.

How can we secure the existence of explanations in terms of shape-gestalt properties for all those phenomenal contrasts that can plausibly be explained by appeal to natural kind properties? The phenomenal contrasts that are cited in support of Siegel's thesis all involve the actualization of a recognitional disposition. We can therefore approach this question by considering different possible ways in which the relevant recognitional dispositions might be actualized. I will consider three different proposals.

First, it is intuitively plausible that we often recognize objects by recognizing their shape-gestalts. In this case, you would recognize something, say, as a pine tree, because you recognize it as having a pine tree shape-gestalt.<sup>33</sup> Now, since we do not have a conclusive scientific account of object-recognition, we do not know whether we ever recognize objects in this way. But if object recognition does work in this way, it follows that any actualization of a recognitional disposition involves the detection of a corresponding shape-gestalt property. And this ensures the existence of an appropriate low-level explanation in these cases of object-recognition.

Second, it is intuitively plausible that we often recognize objects by recognizing a single low-level property or a set of low-level properties.<sup>34</sup> In such cases, one would recognize, say, a familiar person in the dark by his outline shape or a certain bird by its color. But, in my view, these situations

<sup>33</sup> Some influential vision scientists hold that the visual system identifies objects based on three-dimensional shape-representations. This would lend support to the claim that we sometimes recognize objects by recognizing their shape-gestalts. Marr (1982) and Biederman (1987) both have suggested that object recognition is based on three-dimensional shape-representations. Both construed these shape-representations as arrangements of volume-based primitives. Whereas Marr used generalized cylinders as his primitives, Biederman used so-called *geons*. There are many alternative theories of object recognition that are not based on volume-based primitives. One, non-exclusive, alternative is that object-recognition is mediated through representations of two-dimensional object views. For such an account, see Bulthöff and Edelman (1992) and Bulthöff and Wallis (2002).

<sup>34</sup> One low-level property that has received much attention in the empirical literature is motion. See, for example, Johansson (1973) and Mather and Murdoch (1994).



do not generate phenomenal contrasts that can serve as explananda in the first step of the method of phenomenal contrast and, thus, do not need to be explained by the shape-gestalt proposal. The reason for this is that there is no clear difference in sensory phenomenal character between two visual experiences that attribute identical low-level properties, other than shape-gestalt properties, and different natural kind properties to their respective objects. In order to illustrate this, I will now consider two such contrast scenarios.

My first scenario consists of two visual experiences of a red cardinal. Let us say that you know that cardinals are the only red birds in your neighborhood. But you are not able to distinguish cardinals from other red birds by looking, that is, you have not yet acquired the relevant recognitional disposition. As you are walking through the forest, you see a small red dot far off in one of the trees. You cannot make out its shape, but you recognize a number of low-level properties, such as its red color, its location, and its relative size. From these low-level properties together with your background knowledge, you infer that you are looking at a red cardinal. Subsequently, you acquire the ability to recognize cardinals visually. Again, as you are walking through the forest, you see a red dot far off in one of the trees. Suppose also that, in this context, recognizing the object's color, location, and relative size suffices to actualize your recognitional disposition so that you see the object as a cardinal. In this scenario, it seems impossible to tell with any degree of confidence that the two experiences differ in sensory phenomenal character. If this is true, the two experiences of the cardinal do not provide a contrast in sensory phenomenal character that can serve as an explanandum in the first step of the method of phenomenal contrast and, thus, do not need to be explained by the proponent of the low-level view.

In response, the proponent of the high-level view could concede that scenarios of the kind described in the previous paragraph may be conceivable, but deny that they are actually possible. I simply stipulated that, in the second experience, the recognitional disposition is activated by low-level properties, such as color, location, and relative size. But activation of the relevant recognitional disposition might require the experience of other low-level properties. So what we need is a case in which it is more plausible that the recognitional disposition is actualized during the experience. The following scenario is more likely to satisfy this condition.

This scenario consists again of two visual experiences of a red cardinal. The first experience takes place before you acquire the ability to recognize cardinals. You look at a red cardinal that slowly disappears behind a screen and then stops moving. This screen has a small hole in it so that you see only one small part of the cardinal through it. Let us say that you just see a red dot without recognizing any details, such as individual feathers or their texture. The second experience is similar but takes place after you have acquired the disposition to recognize cardinals visually. In this scenario, it is plausible that, in the first experience, you see the dot through the hole



## CAN WE SEE NATURAL KIND PROPERTIES?

as part of a red bird, and, in the second experience, you see the dot as part of a red cardinal.<sup>35</sup> The two experiences attribute identical low-level properties and different natural kind properties to the cardinal. And, yet, it still seems implausible that you could tell with any degree of confidence that the two experiences differ in sensory phenomenal character. Again, the two experiences do not provide a suitable contrast in sensory phenomenal character for the first step of the method of phenomenal contrast.

The proponent of the high-level view could still deny that such scenarios involve the actualization of the relevant recognitional dispositions. Moreover, she could point out that one can accept my conclusions only if one shares my intuitions about the sensory phenomenal character of these imagined scenarios, which are contentious. I admit that both objections are viable. But, at this point in the dialectic, it does not actually matter whether my examples are conclusive. As long as the proponent of the high-level view does not show that putative cases in which we recognize an object by recognizing one or more low-level properties provide clear contrasts in sensory phenomenal character that can serve as explananda in the first step of the method of phenomenal contrast, the shape-gestalt proposal is not threatened.

Third, the proponent of the high-level view could argue that, at least in some cases, object recognition is based on the recognition of natural kind properties. In these cases, recognizing an object as, say, a pine tree, would be nothing other than seeing it as instantiating the natural kind property *being a pine tree*.<sup>36</sup>

I admit that this is a plausible account of object recognition.<sup>37</sup> But this account leads to a problem for the high-level view. Consider again the two experiences of the cardinal behind the screen. According to the high-level view, both experiences would attribute different natural kind properties to the cardinal, namely the property *being a bird* and the property *being a cardinal*. Since, on the high-level view, these properties enter into the sensory phenomenal contents of the two visual experiences, these experiences should clearly differ in their sensory phenomenal character.<sup>38</sup> But, as we have seen, this does not seem to be true. Thus, in order to raise a problem for the shape-gestalt proposal, the proponent of the high-level view would have to explain the absence of a clear difference in sensory phenomenal

<sup>35</sup> One might admit this, but still object that this is not a genuine example of a case where object recognition is necessarily based on the recognition of low-level properties other than shape-gestalt properties. In response, I want to point out that even if the recognitional disposition is originally activated by the viewer's recognition of the cardinal's shape-gestalt, it remains activated after the bird has disappeared behind the screen because the viewer sees a red dot. This might be enough to establish my point.

<sup>36</sup> Siegel indicates that the truth of the high-level view would imply this third notion of object recognition when she writes: "Part of what's at stake about Thesis K [her label for the high-level view] is whether visual experience is an input to such processes of recognition or an output" [Siegel, 2006: 499].

<sup>37</sup> A proposal along these lines has been developed by Prinz. See: Prinz (2006).

<sup>38</sup> The salamander example to be presented later in this paper reinforces this conclusion.



character in the cardinal cases and also present appropriate contrast scenarios that involve a clear difference in sensory phenomenal character.<sup>39</sup>

I have argued that we can provide explanations in terms of shape-gestalt properties for all those phenomenal contrasts that can plausibly be explained in terms of natural kind properties. The proponent of the high-level view could still object that the former explanations do not account for the phenomenal contrast as well as the latter, that is, that they are not phenomenally adequate.<sup>40</sup> However, in order to substantiate this objection, she would have to present contrast scenarios consisting of two visual experiences that satisfy the following conditions: (i) the two experiences attribute identical low-level properties, *including* shape-gestalt properties, to their respective objects, (ii) the two experiences attribute different natural kind properties to their respective objects, and (iii) the two experiences clearly differ in sensory phenomenal character. Yet, I do not think that scenarios that satisfy conditions (i) and (ii) also satisfy condition (iii). As a consequence, we have no reason to believe that explanations in terms of shape-gestalts are phenomenally inadequate. I will first illustrate this with a scenario that is based on an actual case and then argue that this also holds for all possible scenarios that could plausibly be explained in terms of natural kind properties.<sup>41</sup>

The following scenario is based on an actual case of cryptic speciation.<sup>42</sup> Consider a zoologist who is an expert on salamanders. Up to a certain time, she believed that certain salamanders that occur in eastern Mexi-

<sup>39</sup> One might argue that the difference in sensory phenomenal character between seeing something as a bird and seeing something as a cardinal is just too small to be detected because the concepts are closely related to each other. But I do not think that this is a satisfactory response. The reason for this is that one can design scenarios in which the properties attributed by the experiences differ as much as possible.

<sup>40</sup> In fact, Siegel's first argument against the shape-gestalt proposal indicates that she thinks that appeal to shape-gestalt properties does not suffice to explain the phenomenal difference between  $V_1$  and  $V_2$ .

<sup>41</sup> I proceed here by generalization from scenarios based on actual cases. An alternative procedure would be to use twin-Earth scenarios. One could argue as follows. Consider the following twin-Earth scenario. You look at an object here on Earth and see it as being a pine tree. Simultaneously, your twin on twin-Earth looks at an object and sees it as being a twin-pine. A twin-pine looks just like a pine, but has a different genetic make up. That's the only difference between Earth and twin-Earth. In this case, it is implausible that your experience and your twin's experience differ in sensory phenomenal character. Moreover, this seems to be generally true because we can imagine analogous scenarios for any natural kind property. However, twin-Earth scenarios incur a number of problems. For example, twin-Earth scenarios would fail if phenomenal character would supervene on wide content. Various authors have used twin-Earth scenarios to argue for the low-level view, see for example, Price (2006), Brogaard (2013), Prinz (2013). For some problems with these scenarios: see, Siegel (2013) and Logue (2013).

<sup>42</sup> Cryptic speciation is a biological process that results in a group of species (which, by definition, cannot interbreed) that contain individuals that are morphologically identical to each other but belong to different species. For the following example, see, for example, Parra-Olea and Wake (2001). It has recently been argued that there might be different species of giraffes. See, for example, Brown, et al (2007). For evidence of cryptic speciation in maritime organisms, see Cornils and Held (2014).





## CAN WE SEE NATURAL KIND PROPERTIES?

co belong to the same species as certain identical looking salamanders that occur in southern Mexico. DNA testing then showed that the salamanders belong to two different species. Now suppose that the zoologist looks at a terrarium with two compartments. The salamander in the left compartment belongs to the species occurring in eastern Mexico and the salamander in the right compartment belongs to the species occurring in southern Mexico. Suppose also that when the zoologist looks at the two salamanders, the respective recognitional dispositions are actualized. I think that this scenario plausibly satisfies the first two conditions outlined in the previous paragraph. It involves two visual experiences that attribute identical low-level properties, including shape-gestalts, and different natural kind properties to their respective objects. And, yet, it is implausible that the third condition is also satisfied, that is, it is implausible that the two experiences clearly differ in sensory phenomenal character.

I think that it is plausible that the same is true for all possible contrast scenarios that could be explained in terms of natural kind properties. My argument here is simple. We can certainly imagine scenarios of the kind described in the previous paragraph for every natural kind property. And when we do so, we do not notice a clear phenomenal difference between the two imagined visual experiences. Now, there are many cases of cryptic speciation, and, hence, many scenarios of the kind described in the previous paragraph. Given this, it is plausible to assume that imaginability here is evidence for possibility.<sup>43</sup> If this is correct, these kinds of scenarios satisfy conditions (i) and (ii), but fail to satisfy condition (iii). Thus, these scenarios do not present counterexamples to the phenomenal adequacy of the shape-gestalt proposal. We can conclude that the shape-gestalt proposal explains the relevant phenomenal contrasts just as well as appeal to natural kind properties.

One might try to resist my conclusion for a number of reasons. First, one could point out that, like the cardinal scenario, the salamander scenario relies on possibly unreliable intuitions. This might be true. But, again, at this point in the dialectic, this puts the onus on the proponent of the high-level view to provide a phenomenal contrast scenario that satisfies conditions (i) to (iii). Lacking such a scenario, we have no reason for thinking that the shape-gestalt proposal is phenomenally inadequate.

Second, the proponent of the high-level view could point out that the two visual experiences in the salamander case represent the two different natural kind properties by identical sensory phenomenal characters.<sup>44</sup> Thus, even if my description of this scenario is correct, it does not follow that these properties do not enter into the sensory phenomenal contents of

<sup>43</sup> Note that this is not the kind of imaginability involved in twin-Earth scenarios. In contrast to twin-Earth scenarios, I imagine scenarios based on experiences had by the same subject in the actual world.

<sup>44</sup> Siegel brings up this option in her response to Prinz. See: Siegel (2013: 851).



the two experiences. In order to show this, one would have to prove that there is no difference in sensory phenomenal character between a visual experience of the salamander that lacks the recognitional element and one that represents it as a salamander.

I do not think that this objection is valid, however. According to the proponent of the high-level view, different natural kind properties are typically represented by different sensory phenomenal contents. We would thus need a reason for why this is different in the salamander scenario. Moreover, we can construe scenarios in which one of the salamanders turns out to be a bizarre atmospheric disturbance, a strange type of moving plant, a strange looking elephant, or whatever. Thus, even if the proponent of the high-level view could give us an explanation in the original salamander case, it is implausible that this explanation would cover all the other cases.

## 6. Conclusion

My goal in this paper was to show that the shape-gestalt proposal provides a viable low-level alternative to Siegel's high-level view. I proceeded roughly in three steps. I first showed that Siegel's arguments against the shape-gestalt proposal are not conclusive. I then developed the shape-gestalt proposal in greater detail. Finally, I argued that the shape-gestalt proposal allows us to explain the relevant phenomenal contrasts just as well as Siegel's high-level view. The argument for this latter claim proceeded in two steps. I first showed that we can provide explanations in terms of shape-gestalt properties for all those phenomenal contrasts that can plausibly be explained in terms of natural kind properties. I then argued that explanations in terms of shape-gestalt properties are phenomenally adequate. My argument does not show that Siegel's proposal is false. But, if correct, it establishes the existence of a plausible low-level alternative.

## References

- Austin, J. L. 1962. *Sense and sensibilia*. Oxford: Clarendon Press.
- Bayne, T. (2009). Perception and the reach of phenomenal content. *The Philosophical Quarterly*, 59, 386–404.
- Bayne, T. and Montague, M. (2011). Cognitive phenomenology: an introduction. In (Eds.) T. Bayne and M. Montague, *Cognitive Phenomenology*. Oxford: Oxford University Press, 1–34.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Brewer, B. (2006). Perception and content. *European Journal of Philosophy* 14, 165–181.



## CAN WE SEE NATURAL KIND PROPERTIES?

Brogaard, B. (2013). Do we perceive natural kind properties? *Philosophical Studies*, 162, 35–42.

Brown, D.M. (2007). Extensive population genetic structure in the giraffe. *BMC Biology* 2007, 5, 57.

Bulthöff, H.H. and Edelman, S. (1992). Psychophysical support for a two-dimensional view-interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the USA*, 89, 60–64.

Bulthöff, H.H. and Wallis, S. (2002). Learning to recognize objects. In M. Fahle and T. Poggio (Eds.), *Perceptual learning*. MIT Press.

Byrne, A. (2009). Experience and content. *The Philosophical Quarterly*, 59, 429–451.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Carruthers, P. and Veillet, B. (2012). The case against cognitive phenomenology. In T. Bayne and M. Montague, *Cognitive Phenomenology* (pp. 35–56), Oxford: Oxford University Press.

Cornils, A. and Held, Ch. (2014). Evidence of cryptic and pseudocryptic speciation in the *Paracalanus parvus* species complex (Crustacea, Copepoda, Calanoida). *Frontiers in Zoology* 2014, 11:19.

Crane, T. (2006). Is there a perceptual relation? *Perceptual Experience*. Ed. T. S. Gendler, J. Hawthorne. Oxford: Oxford University Press, 126–146.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge: MIT Press.

Fish, W. (2009). *Perception, hallucination, and illusion*. Oxford: Oxford University Press.

Gauthier, I. and Tarr, M.J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. *Vision Research*, 37, 1673–82.

Gombrich, E.H. (1960). *Art and illusion*. Princeton: Princeton University Press.

Gopnik, A. and Rosati, A. (2001). Duck or rabbit? Reversing ambiguous figures and understanding ambiguous representations. *Developmental Science*, 4, 175–183.

Horgan, T. (2011). From agentive phenomenology to cognitive phenomenology. In T. Bayne and M. Montague, *Cognitive phenomenology* (pp. 57–78), Oxford: Oxford University Press.

Horgan, T. and Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In David J. Chalmers (Ed.), *Philosophy of mind: classical and contemporary readings*. Oxford University Press.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14, 201–211.

Johnston, M. (2004). The obscure object of hallucination. *Philosophical Studies*, 120, 113–183.

Johnston, M. (2006). Better than mere knowledge? The function of sensory awareness. *Perceptual Experience*. Ed. T. S. Gendler, J. Hawthorne. Oxford: Oxford University Press, pp. 260–290.

Kimchi, R., Behrman, M., and Olson, C.R. (2003). *Perceptual organization in vision: Behavioral and neural perspectives*. Lawrence Erlbaum: London

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*, Cambridge, MA: MIT Press.

Kriegel, U. (2007). The phenomenologically manifest. *Phenomenology and the Cognitive Sciences*, 6, 115–136.



- Lamme, V. A. F. and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23, 571–79.
- Logue, H. (2013). Visual experience of natural kin properties: is there any fact of the matter. *Philosophical Studies*, 162, 1–12.
- Lopes, D. (1996). *Understanding pictures*. Cambridge: Cambridge University Press, 1996.
- Lycan, W. G. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.
- Lyons, J. (2005). Perceptual belief and non-experiential looks. In J. Hawthorne (Ed.), *Philosophical Perspectives*, 19: *Epistemology* (pp. 237–256), Malden: Blackwell.
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco, CA: W.H. Freeman.
- Martin, M. G. F. (2004). The limits of self-awareness. *Philosophical Studies*, 120, 37–89.
- Masrour, F. (2011). Is perceptual phenomenology thin? *Philosophy and Phenomenological Research*, 83(2), 366–397.
- Mather, G. and Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London (B)*, 258, 273–278.
- McDowell, J. (1994). *Mind and world*. Cambridge, MA: Harvard University Press.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–50.
- Nanay, B. (2011). Do we perceive apples as edible? *Pacific Philosophical Quarterly*, 92, 305–322
- Nanay, B. (forthcoming). Action-oriented perception. *European Journal of Philosophy*.
- Noë, A. 2004. *Action in perception*. Cambridge, MA: The MIT Press.
- Parra-Olea, G. and D.B. Wake. 2001. Extreme morphological and ecological homoplasy in tropical salamanders. *Proceedings of the National Academy of Sciences*, 98, 888–7891.
- Pautz, A. (2009). What are the contents of experiences? *The Philosophical Quarterly*, 59, 483–507.
- Pautz, A. (2010). Why explain visual experience in terms of content?. In B. Nanay (Ed.), *Perceiving the world*, (pp. 254–309), Oxford: Oxford University Press.
- Peacocke, C. (1992). *A Study of concepts*. Oxford: Clarendon Press.
- Pitt, D. (2004). The phenomenology of cognition or what it is like to think that P. *Philosophy and Phenomenological Research*, 69, 1–36.
- Price, R. (2009). Aspect-switching and visual phenomenal character. *The Philosophical Quarterly*, 59, 508–518.
- Prinz, J. (2006). Beyond appearances: The content of sensation and perception. In T. Gendler and J. Hawthorne (Eds.), *Perceptual experience*, (pp. 434–460), Oxford: Oxford University Press.
- Prinz, J. (2013). Siegel's get rich quick scheme. *Philosophical Studies*, 163, 827–835.
- Reiland, I. (2014). On experiencing high-level properties. *American Philosophical Quarterly*, 51, 2014, 177–187



- Rock, I. and Mitchener, K. (1992). Further evidence of failure of reversal of ambiguous figures by uninformed subjects. *Perception*, 21, 39–45.
- Rock, I, Gopnik, A., and Hall, S. (1994). Do young children reverse ambiguous figures? *Perception*, 23, 635–644.
- Russell, B. (1912). *The problems of philosophy*. Oxford: Oxford University Press.
- Schellenberg, S. (2011). Perceptual content defended. *Nous*, 45(4), 714–750.
- Schier, F. (1986). *Deeper into pictures*. Cambridge: Cambridge University Press.
- Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, 117, 245–273.
- Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Siegel, S. (2006). Which properties are represented in perception? In T. Gendler and J. Hawthorne (Eds.), *Perceptual Experience* (pp. 481–503), Oxford: Oxford University Press.
- Siegel, S. (2007). Can we discover the contents of experience. *Southern Journal of Philosophy*, 45 (S1), 127–142.
- Siegel, S. (2010). *The contents of visual experience*. Oxford: Oxford University Press.
- Siegel, S. (2013). Replies to Campbell, Prinz, and Tavis. *Philosophical Studies*, 163, 847–865.
- Siewert, C. (1998). *The significance of consciousness*. Princeton, N.J.: Princeton University Press.
- Siewert, S (2011). Phenomenal thought. In T. Bayne and M. Montague, *Cognitive Phenomenology* (pp. 236–267), Oxford: Oxford University Press.
- Smith, A.D. (2002). *The problem of perception*. Cambridge: Harvard University Press.
- Strawson, G. (1994). *Mental reality*. Cambridge, MA: MIT Press.
- Strawson, G. (2011). Cognitive phenomenology: Real life. In T. Bayne and M. Montague, *Cognitive Phenomenology* (pp. 285–325), Oxford: Oxford University Press.
- Travis, Charles (2004). The silence of the senses. *Mind*, 113, 57–94.
- Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: MIT Press.



## CAN ANOSOGNOSIA VINDICATE TRADITIONALISM ABOUT SELF-DECEPTION?

**José Eduardo Porcher** –  
PhD. Federal University  
of Rio Grande do Sul,  
Porto Alegre, Brazil.  
E-mail:



The traditional conception of self-deception takes it for an intrapersonal form of interpersonal deception. However, since the same subject is at the same time deceived and deceiving, this means attributing the agent a pair of contradictory beliefs. In the course of defending a deflationary conception of self-deception, Mele (1997) has challenged traditionalists to present convincing evidence that there are cases of self-deception in which what he calls the dual belief-requirement is satisfied. Levy (2009) has responded to this challenge affirming that there is at least one real case of self-deception that meets this requirement, namely, that of anosognosia. In this family of conditions, the patient apparently believes that there is nothing wrong with her while, at the same time, providing behavioral cues that indicate that the patient is somehow aware of his disease. If Levy is right, then traditionalism about self-deception could be vindicated, after having been widely abandoned due to its need to postulate exotic mental processes in order to make sense of the attribution of contradictory beliefs. In this paper, I assess whether Levy's response to Mele's challenge is successful by analyzing his interpretation of the empirical evidence to which he appeals. Finally, I attack the cogency of the underlying commitments about the nature of folk psychology to which one is required to defer in order to draw from conflicting evidence the attribution of contradictory beliefs.

**Key words:** self-deception. anosognosia. belief. contradictory beliefs. folk psychology.

### Introduction

Self-deception is a psychological phenomenon with which every human being is familiar. Since the definition of self-deception is precisely one of the points of contention in the philosophical literature, one might as well begin by first determining the reference of the concept. There is no shortage of examples, but maybe it will help if we confine ourselves to those which are relatable and less intricate. Consider the following representative case:

Ann is dying of cancer and is aware of many facts, such as her long, steady decline, pointing to this outcome, though no one has told her that her case is terminal and she has avoided letting her doctor give her a prognosis. Suppose further that she talks of recovery and discusses her various plans for the long future ... the facts pointing to her death are not unmistakably prominent and her talk of recovery is apparently sincere ... she has better than average medical knowledge ... [but] (among other things) ... her talk of recovery lacks full conviction (or exhibits too much apparent conviction), and ... is often followed by depression or anxiety [Audi, 1982:134].

Despite our familiarity with this type of situation, and although the folk-psychological concept of self-deception (or its equivalents, such as fooling oneself, lying to oneself, etc.) is used by us every day, we have yet to come up with a consensual cha-



## CAN ANOSOGNOSIA VINDICATE TRADITIONALISM...

racterization of the phenomenon. This is due to the fact that, though the phenomenon is straightforward enough, there seems to be something amiss with the very idea of deceiving oneself. This puzzle gave rise to a specialized debate with the translation of Jean-Paul Sartre's *Being and Nothingness* into English. In his discussion of "bad faith," Sartre recognizes and elaborates on the apparent contradiction involved in self-deception:

I must know in my capacity as deceiver the truth which is hidden from me in my capacity as the one deceived. Better yet, I must know the truth very exactly in order to conceal it more carefully—and this not at two different moments, which at a pinch would allow us to reestablish a semblance of duality—but in the unitary structure of a single project (1949/1957, p. 49):

So the very word 'self-deception' carries with it an air of impossibility if we take it to mean exactly what it seems to mean. On close inspection, two puzzles arise from a literal interpretation of the word, each of which is derived from one of two lexical assumptions:

1. By definition, person *A* deceives person *B* (where *B* may or may not be the same person as *A*) into believing that *p* only if *A* knows, or at least believes truly, that not-*p* and causes *B* to believe that *p*.
2. By definition, deceiving is an intentional activity: nonintentional deceiving is conceptually impossible [Mele, 2001: 6].

The first puzzle, often called the *static puzzle*, concerns the possibility of the mental state that the first lexical assumption entails, namely, the possession of contradictory beliefs. The second puzzle, often called the *dynamic puzzle*, concerns the possibility of the project that the second lexical assumption entails, namely, intentionally keeping from oneself something that one believes. For the remainder of this paper, I will deal with the first of these difficulties.

There have been two main proposed approaches to the static puzzle. The first one takes 'deception' literally, declaring self-deception to be a form of *intrapersonal* deception. I will refer to this as *traditionalism about self-deception*. Insofar as traditionalism entails that the belief that *p* and the belief that not-*p* coexist in the mind of the self-deceived, the static puzzle must be solved. The solutions proposed by traditionalists such as David Pears (1984) and Donald Davidson (1985) rest on the Freudian idea that the best way to account for the phenomenon is to somehow *split* the person. Pears's proposed solution is the most radical of these:

[There is a] subsystem ... built around the nucleus of the wish for the irrational belief and it is organized like a person. Although it is a separate center of agency within the whole person, it is, from its own point of view, entirely rational. It wants the main system to form the irrational belief and it is aware that it will not form it, if the cautionary belief [i.e., the belief that it would be irrational to form the desired



belief] is allowed to intervene. So with perfect rationality it stops its intervention (1984: 87),

It is easy to see how this would solve the pending difficulties. Pears converts the problematic characterization ‘*A* deceives *A*’ that resulted from a literal reading into the non-problematic ‘*A* deceives *B*,’ where *A* and *B* are different subsystems of agency within a reasonably unified system, namely, the person. Because the roles of deceiver and deceived are played by different centers of agency, the aura of paradox disappears. However, this sort of explanation faces its own difficulties. As Mark Johnston (1988: 64) has observed, the traditionalist view replaces a contradictory description of the self-deceiver with one that raises its own host of problems: how can the deceiving subsystem have the capacities to perpetrate the deception? Why should the deceiving subsystem be interested in the deception? Does it suppose that it knows what it is best for the deceived system to believe? Importantly, as Tamar Gendler observes, ‘If one of the subpersons (truly) believes that *p* and does not believe that not-*p*, and if that subperson is bothered by this and wishes it were not the case, why would she be find it psychologically fruitful intentionally to bring *someone else* to believe that the opposite?’ (2007: 235).

On the other hand, Davidson proposes a functional division that falls short of literally splitting the person into separate centers of agency. His view is that all that is needed is a boundary between conflicting attitudes: there would be no problem in believing contradictory propositions if they didn’t come in contact with each other. Davidson claims that it is the drawing of such a boundary between our inconsistent beliefs which constitutes the irrational step involved in self-deception, and that this step is assisted by the nonobservance of Carl Hempel’s and Rudolf Carnap’s requirement of total evidence for inductive reasoning—a normative principle that enjoins us to give credence to the hypothesis most highly supported by all available relevant evidence when choosing among a set of mutually exclusive hypotheses. Davidson proposes the following conditions under which an agent *A* is self-deceived with respect to a proposition *p*:

*A* has evidence on the basis of which he believes that *p* is more apt to be true than its negation; the thought that *p*, or the thought that he ought rationally to believe *p*, motivates *A* to act in such a way as to cause himself to believe the negation of *p* (1985: 88).

Nevertheless, even though Davidson’s is a more economical approach to the characterization of the self-deceived than Pears’s, both approaches share the assumption that ‘self-deception’ should be understood literally. However, the ‘deception’ in self-deception may alternatively be understood as nothing more than a *metaphor* (an observation that originated as early as Canfield and Gustafson, 1962). For proponents of this view, self-deception is not to be understood as a reflexive form of deception, in





much the same way that self-teaching is not understood as a reflexive form of teaching.

The second approach to the static puzzle involves understanding self-deception to be simply a form of motivated irrationality. I will refer to this as *deflationism about self-deception*. Insofar as deflationism maintains that only the false or unwarranted belief exists in the mind of the self-deceived, the static puzzle is apparently bypassed. Deflationists are, however, left with the task of explaining what exactly the mental states involved in self-deception are, and how they are formed and maintained. In other words, their task is explaining what the metaphor stands for. Alfred Mele, the main proponent of deflationism, presents the following set of jointly sufficient conditions for an agent *S* to be self-deceived in acquiring a belief that *p*:

1. The belief that *p* which *S* acquires is false.
2. *S* treats data relevant, or at least seemingly relevant, to the truth value of *p* in a motivationally biased way.
3. This biased treatment is a nondeviant cause of *S*'s acquiring the belief that *p*.
4. The body of data possessed by *S* at the time provides greater warrant for not-*p* than for *p* (2001: 51).

Although it is certainly not a given that these conditions are indeed jointly sufficient for self-deception to obtain, deflationism has the virtue of at once avoiding the static puzzle and presenting a more parsimonious characterization of the self-deceived.

### Mele's challenge and Levy's response

Having advanced the most parsimonious theory of self-deception to date, Mele (1997) challenged traditionalists to provide convincing evidence of the existence of instances of self-deception that satisfy what he calls the *dual-belief requirement*, namely, the requirement that self-deceivers (at some point) simultaneously believe that *p* and believe that not-*p*. Recently, Neil Levy (2009) has responded to Mele's challenge with an attempted demonstration of the satisfaction of the dual-belief requirement. While recognizing that—given that Mele offers merely sufficient (and not necessary) conditions—meeting the challenge would not falsify the deflationary account, Levy points out that, if he is successful, he will have at least undermined the argument from parsimony in favor of deflationism and have strengthened traditionalism insofar as there would not be a need to postulate *ad hoc* exotic mental processes such as the ones invoked by Pears (and, to some extent, Davidson).<sup>1</sup>

<sup>1</sup> However, Levy doesn't aim at vindicating traditionalism entirely, as he doesn't defend intentionalism. What he does intend to do is to 'show that the static puzzle must be solved rather than simply dismissed' (2009: 230).



How may one go about demonstrating that at least some cases of self-deception involve the simultaneous possession of contradictory beliefs? As Levy notes, ‘we cannot directly inspect the mind of the self-deceived in order to discover that there are cases where the dual-belief requirement is satisfied’ (2009: 230). What one can do is to analyze the evidence available for the attribution of contradictory beliefs to subjects in the grip of self-deception. Levy makes it clear that he thinks that ordinary self-deception may be properly characterized without the attribution of contradictory beliefs, but suggests he is on to at least one case of extraordinary self-deception that does satisfy the dual-belief requirement, making it necessary at least to amend the deflationary account. The case at hand is that of anosognosia.

Anosognosia is a neurological condition characterized by the lack of awareness or the underestimation of a specific deficit in sensory, perceptual, motor, affective or cognitive functioning, brought about by brain dysfunction (Bisiach and Geminiani, 1991; Prigatano, 2010). In the following discussion I will follow Levy and use the term anosognosia to refer to the unawareness of a particular motor deficit, hemiplegia – paralysis of the arm, leg, and/or trunk on the same side of the body – as anosognosia ‘has been studied mainly in stroke hemiplegic patients, who may report no deficit, overestimate their abilities or deny that they are unable to move a paretic limb’ (Orfei et al., 2007: 3075). The vast majority of cases of anosognosia for hemiplegia follow right-brain injury and thus it is the left side of the patient’s body that is usually paralyzed.

What leads Levy to think the case of anosognosia has any significance for the debate on the proper characterization of self-deception is the fact that numerous descriptions and interpretations of the denial, rationalization, and confabulation involved in cases of anosognosia invite comparison with the motivated phenomenon of self-deception. The main source of case reports from which Levy draws comes from Vilayanur S. Ramachandran’s studies (1995; 1996). To understand why, according to Levy, it is tempting to see anosognosia as a case of extreme self-deception, let us look at two representative cases. Here is a conversation between one of Ramachandran’s patients and himself (Ramachandran and Blakeslee, 1998: 130):

“Esmerelda, how are you doing?”

“I’m fine?”

“Can you walk?”

“Yes.”

“Can you use your arms?”

“Yes.”

“Can you use your right arm?”

“Yes”.

“Can you use your left arm?”?

“Yes, I can use my left arm.”



## CAN ANOSOGNOSIA VINDICATE TRADITIONALISM...

“Can you point to me with your right hand?”  
She pointed straight at me with her good right hand.  
“Can you point to me with your left?”?  
Her left hand lay motionless in front of her?  
“Esmerelda, are you pointing?”  
“I have severe arthritis in my shoulder; you know that, doctor. It hurts.  
I can’t move my arm now.”  
On other occasions she employed other excuses:?  
“Well, I’ve never been very ambidextrous, doctor.”

Patients of anosognosia are prone to rationalization, making up excuses in much the same way that ordinary people do when self-deceived. This invites not only the comparison with self-deception, but the hypothesis that the neurological mechanisms involved in both phenomena are the same or, at least, somehow related (Hirstein, 2005).<sup>2</sup> Some cases of anosognosia, however, are even more stark and invite the ascription of full-blown delusion.<sup>3</sup> Here is another conversation between one of Ramachandran’s patients and himself (Ramachandran and Blakeslee, 1998: 129):

“Mrs. Dodds, can you touch my nose with your right hand?”  
She did so with no trouble.  
“Can you touch my nose with your left hand?”  
Her hand lay paralyzed in front of her?  
“Mrs. Dodds, are you touching my nose?”  
“Yes, of course I’m touching your nose.”  
“Can you actually see yourself touching my nose?”?  
“Yes, I can see it. It’s less than an inch from your face.”

At this point Mrs. Dodds produced a frank confabulation, almost a hallucination, that her finger was nearly touching my nose. Her vision was fine. She could see her arm perfectly clearly, yet she was insisting that she could see the arm move? I decided to ask just one more question.

“Mrs. Dodds, can you clap?”  
With resigned patience she said, “Of course I can clap.”  
“Will you clap for me?”?

Mrs. Dodds glanced up at me and proceeded to make clapping movements with her right hand, as if clapping with an imaginary hand near the midline?

“Are you clapping?”?  
“Yes, I’m clapping,” she replied.

As Ramachandran notes, this patient’s confabulation is on the extreme end of the scale, as it is much more usual for patients with anosognosia to conjure up inane excuses or rationalizations why their left arms do not mo-

<sup>2</sup> If this were true, it would mean that self-deception might at last begin to be understood at a lower level of description—namely, that of neurobiology. This alone makes the comparison worth investigating.

<sup>3</sup> This raises the interesting question of where self-deception ends and clinical delusion begins. Cf. Bortolotti and Mameli (2012).



ve when they are asked to demonstrate the use of that arm. Hence, this case is an extraordinary case of anosognosia and ordinary cases of anosognosia still invite comparison with everyday self-deception, albeit in an exaggerated form. Levy uses this as a starting point to explore the possibility that anosognosia might be evidence against deflationism about self-deception. His overarching argument starts from the premises that (1) some cases of anosognosia are cases of self-deception, and that (2) these cases warrant the ascription of contradictory beliefs, concluding that there is at least one kind of case of self-deception which warrants the ascription of contradictory beliefs. As a corollary of Levy's argument, deflationism about self-deception turns out to be either incorrect or at least incomplete.

Whether or not Levy's first premise is true depends, on the one hand, on an account of self-deception, and on the other hand, on an account of anosognosia. Levy gives the following set of conditions for a case of anosognosia to be a case of self-deception (2009: 234):

1. Subjects believe that their limb is healthy.
2. Nevertheless they also have the simultaneous belief (or strong suspicion) that their limb is significantly impaired and they are profoundly disturbed by this belief (suspicion).
3. Condition #1 is satisfied because condition #2 is satisfied; that is, subjects are motivated to form the belief that their limb is healthy because they have the concurrent belief (suspicion) that it is significantly impaired and they are disturbed by this belief (suspicion).

Note that Levy's characterization of self-deception is somewhat imprecise, since he attributes to the self-deceived the desired, but unwarranted belief, while leaving it open if the attitude of the self-deceived toward the undesired state of affairs is one of belief or suspicion. (I will return to this point.) However, from the point of view of an account of anosognosia, the main point of contention is his third condition, which concerns the motivational origin of anosognosic beliefs. Of course, Levy is not stating that *all* anosognosic beliefs are the product of the subject's desires, but it is still a matter of controversy that there are any motivationally formed anosognosic beliefs (cf. Aimola Davies et al., 2009). Still, since the focus of this investigation is not whether or not it is true that some cases of anosognosia are cases of self-deception, I will assume, for the sake of the argument, that the first premise of Levy's argument is true and turn now to his second premise—the claim that anosognosia warrants the ascription of contradictory beliefs. To support the second premise of his argument, Levy intends to demonstrate that there is evidence to attribute both the desired belief that one is not paralyzed and the undesired belief that one is paralyzed. In the next section, I present and assess this evidence.



## Analyzing the evidence

As we have seen, anosognosic patients deny being paralyzed with apparent sincerity. Importantly, this conviction has been empirically tested. In order to answer the question ‘Is there tacit knowledge of paralysis in anosognosia?’, Ramachandran devised an experiment in which patients were given the choice of completing either a bimanual or a unimanual task for larger and smaller prizes, respectively. He administered the trials to three different patients a total of 19 times. Patients almost always chose the bimanual task, leading Ramachandran to assert that ‘the patients either have no “tacit knowledge” of their paralysis or, even if they do, they cannot access this knowledge when choosing between a unimanual vs. bimanual task’ (1995: 31). So far, then, the evidence heavily suggests that anosognosic patients believe that they are not paralyzed, and that they fail to believe that they are paralyzed. Nevertheless, Levy assembles several pieces of evidence which, he purports, while not indisputable on its own, make a strong case for the simultaneous attribution of the undesired belief.

First, Levy argues that there is evidence that the fact of paralysis is represented in the brain of anosognosic patients (2009: 236). This is suggested by studies employing vestibular stimulation through the caloric reflex test—a test of the vestibulo-ocular reflex that involves irrigating cold or warm water into the external auditory canal. This test is hypothesized to arouse the parts of the right hemisphere normally engaged in anomaly detection and attention to the left side of the patient’s personal space and, notably, it has been shown to lead to temporary remission of anosognosia. More importantly, Levy cites the fact that the application of this test to anosognosic patients has showed recognition that they had awareness of their paralysis all along (Cappa et al., 1987). But note that this is not conclusive evidence given that the might be the alleged recognition may well be an effect of hindsight bias. More importantly, however, is the question of how something being ‘represented in the brain’ would equal something being believed. Moreover, Hirstein (2005) suggests that we take anosognosics at their word and that, if their brain can nevertheless represent the damage to their limb—a possibility which is not being rejected here at all—this representation is *subpersonal* and inaccessible to consciousness.

Second, Levy cites indirect evidence that the relevant proposition has (at least) the lowest degree of personal availability to the anosognosic patient (2009: 237). Forced-choice situations have yielded evidence of implicit processing in patients with hemispatial neglect—a condition in which, after damage to one hemisphere of the brain, a deficit in attention to and awareness of one side of personal space is observed. Neglect is defined by the inability of a person to process and perceive stimuli on one side of the body or environment that is not due to a lack of sensation (Unsworth: 2007). Marshall and Halligan (1988) showed a neglect patients drawings of



houses, placed so that the leftmost part of the houses fell in her neglected field. Predictably, the patient reported that the houses looked identical. However, the drawing of one the houses was in flames on its left side, something which the patient could not *consciously* see. Marshall and Halligan then proceeded to ask the patient which house they would prefer to live in and reliably chose the house that was not burning. Note, however, that even if this result were undisputed (Fahle, 2003), Levy doesn't offer grounds for accepting evidence from neglect to apply to anosognosia across the board. Nevertheless, even if we were to accept the evidence of implicit processing from neglect patients to transfer to anosognosia patients, Levy again gives us no reason to infer that that the patient *believes* the relevant proposition from evidence that establishes only (if at all) what he himself terms 'the lowest degree of personal availability'—that of cases of blindsight, in which patients cannot use visual information from their blind field in everyday life, but are able to use it to guess above chance in forced-situations (Weiskrantz, 1986).

Third, Levy claims that there is observational evidence that the explicitly denied knowledge guides some of the behavior of anosognosic patients, including their verbal behavior, and takes this to indicate that such knowledge has a degree of availability somewhat above that of visual information in blindsight (2009: 237). This alleged evidence comes from two cases reported by Ramachandran. The first case (Ramachandran and Blakeslee, 1998: 139) is that of a patient who, after opting for a bimanual task in a forced-choice situation—namely, tying her shoelaces—and failing to complete it, went on to affirm afterwards that she *had* tied her shoelaces. The second case (Ramachandran and Blakeslee, 1998: 150) is that of a patient who affirmed that his left, paralyzed arm was actually *stronger* than his right, healthy arm.

Note that it isn't obvious how these cases represent evidence that these patients's alleged belief that they are paralyzed guides their behavior. To arrive at that conclusion, Levy buys into Ramachandran's interpretation of these cases. According to Ramachandran, both behavioral manifestations are indicative of what Freud called "reaction formation," defined by Levy as 'the expression of a thought antithetical to the denied proposition, which betrays its motivated nature by its very vehemence' (2009: 237). While this is a valid interpretation, it by no means establishes either that anosognosic patients are somehow aware of their condition—since the manifested behavior can also be interpreted as the result of a frank confabulation—or that any motivational component is involved in anosognosia at all. Moreover, note that once more the language employed by Levy warrants questioning why we should get onboard with belief attribution if the evidence warrants but only a degree of availability somewhat above that of visual information in blindsight.

Fourth, Levy argues that there is strong evidence that the denied knowledge is dispositionally available to anosognosic patients (2009: 237).



Once again, Levy turns to Ramachandran, citing the fact that patients can be gently “prodded” into eventually admitting that their arm is not working, weak and even, in some cases, paralyzed, although, as Ramachandran and Blakeslee note, they seem unperturbed by this admission (1998: 149), a qualification that Levy fails to acknowledge. But even if patients could be prodded into admitting that they are paralyzed, it is by no means trivial that we should immediately take this as evidence that knowledge of the paralysis is dispositionally available to the patients rather than, for example, as evidence that the patient has just *acquired* such knowledge. In the absence of reasons why the latter plausible interpretation is incorrect, Levy’s alleged evidence cannot do the required explanatory work.

### The attribution of contradictory beliefs

I claim that the evidence Levy presents for the presence of the undesired belief, as he himself gestures at repeatedly by his choice of words, is not enough to attribute a belief. After presenting the four pieces of evidence for the undesired belief above, Levy concludes that ‘taken together, this evidence seems to constitute a strong case for attributing to anosognosics the belief or the strong suspicion that their limb is significantly impaired’ (2009: 238). I disagree. While what Levy presents as evidence of the presence of the undesired belief is weak on its own since, as we have seen, it establishes but a low degree of availability, it becomes even weaker when we simultaneously have very strong reasons to attribute the desired belief, that is, to take anosognosic patient’s at their word.

Levy’s conclusion that we have enough evidence to attribute to the anosognosic patient the belief that he or she is paralyzed stems from the *isolation* of this evidence from the evidence that he himself points out is enough to warrant the attribution to the anosognosic of the belief that he or she is healthy. He then takes the sum total of the evidence to warrant the attribution of contradictory beliefs to the anosognosic and, given that he understands at least some cases of anosognosia to be cases of self-deception, he purports to have demonstrated that Mele’s dual-belief requirement holds in at least some cases of self-deception. I argue, in turn, that even if it were true that anosognosia is (in some cases, at least) a form of self-deception, Levy would not be right to derive an attribution of contradictory beliefs from evidence that pulls in two opposing directions. This is because, in the absence of an empirical method of direct inspection of the anosognosic patient’s mind (as acknowledged by Levy), *belief is an explanatory, interpretive, and predictive concept*—a tool we use to explain, interpret, and predict behavior. And if we’re presented with conflicting evidence, the right response is not to attribute contradictory beliefs—since these have no explanatory power at all—but to withhold attribution of belief and opt, inste-



ad, for a characterization of the subject's attitudes that doesn't abstract away from the cognitive processes that underlie belief attribution. For the sake of advancing in the search for a scientific theory of the phenomena discussed, we need to recognize that cases such as that of anosognosia, and clinical delusions in general, are often simply not amenable to personal-level description and explanation. When it only pretends to be describing precisely the mental states of those to which we attribute beliefs, 'belief' talk does more harm than good.

## Conclusion

Levy rightly points out that, like self-deception, anosognosia 'presents us with a real-life—indeed, clinically verified—case in which agents claim that  $p$ , while nevertheless giving clear indications that they (at least) strongly suspect that not- $p$ ' (2009: 239). Instead of relying on imaginary or literary cases of self-deception, as almost all authors contributing to the literature have done, Levy has taken an important step in bringing empirical evidence to bear on the discussion. However, I have argued that, notwithstanding his efforts to use anosognosia to demonstrate that Mele's dual-belief requirement for self-deception is met, Levy fails to establish this. Importantly, I claim, on explanatory grounds, that Mele's challenge is not answerable by stipulating that evidence pulling us in the direction of opposite attributions warrants our attributing contradictory beliefs to anosognosic patients. In fact, after Levy's attempted demonstration, we have nothing really new that would warrant the attribution of contradictory beliefs to the self-deceived, assuming that anosognosia really is (sometimes, at least) a form of self-deception. Cases of everyday self-deception *also* push us in opposite directions and give us evidence of doxastic conflict. This is precisely what has fueled years of debate on what the self-deceived *really* believe—if they believe that  $p$  and not- $p$ , only  $p$ , only not- $p$ , or, as I contend, that the question doesn't have to have an answer, since assuming so is an overestimation of the preciseness of our folk-psychological vocabulary. If the evidence of doxastic conflict alone were enough for concluding that the dual-belief requirement is met, then the debate would never even have started.

## References

- Aimola Davies, A.M, Davies, M., Ogden, J.A., Smithson, M, and White, R.C. (2009). Cognitive and Motivational Factors in Anosognosia. In T. Bayne and J. Fernandez, *Delusion and Self-Deception: Affective and Motivational Influences in Belief Formation*. Psychology Press.
- Audi, R. (1982). Self-Deception, Action, and Will. *Erkenntnis* 18: 133–58.





Bisiach, E. and Geminiani, G. (1991). Anosognosia related to hemiplegia and hemianopia. In G.P. Prigatano and D.L. Schacter, eds., *Awareness of Deficits After Brain Injury*. Oxford University Press.

Bortolotti, L. and Mameli, M. (2012). Self-deception, delusion and the boundaries of folk psychology. *Humana.Mente* 20: 203–221.

Cappa, S., Sterzi, R., Vallar, G. and Bisiach E. (1987). Remission of hemineglect and anosognosia during vestibular stimulation. *Neuropsychologia* 25: 775–82.

Canfield, J.V. and Gustafson, D.F. (1962). Self-Deception. *Analysis* 23: 32–36.

Davidson, D. (1985). Deception and Division. In J. Elster, ed., *The Multiple Self*. Cambridge University Press.

Fahle, M. (2003). Failures of visual analysis: scotoma, agnosia, and neglect. In M. Fahle and M. Greenlee, eds. *The Neuropsychology of Vision*. Oxford University Press.

Gendler, T.S. (2007). Self-deception as pretense. *Philosophical Perspectives* 21 (Philosophy of Mind), 231–258.

Hirstein, W. (2005). *Brain Fiction: Self-Deception and the Riddle of Confabulation*. MIT Press.

Johnston, M. (1988). Self-Deception and the Nature of Mind. In B. McLaughlin and A. Rorty, eds., *Perspectives on Self-Deception*. Berkeley: University of California Press.

Levy, N. (2009). Self-Deception Without Thought Experiments. In T. Bayne and J. Fernandez, *Delusion and Self-Deception: Affective and Motivational Influences in Belief Formation*. Psychology Press.

Marshall, J.C. and Halligan, P.W. (1988). Blindsight and insight in visuo-spatial neglect. *Nature* 336: 766–767.

Mele, A.R. (1997). Real Self-Deception. *Behavioral and Brain Sciences* 20: 91–102.

Mele, A.R. (2001). *Self-Deception Unmasked*. Princeton University Press.

Orfei, M.D., Robinson, R.G., Prigatano, G.P., Starkstein, S., Rüşch, N., Bria, P., Caltagirone, C. and Spalletta, G. (2007). Anosognosia for hemiplegia after stroke is a multifaceted phenomenon: a systematic review of the literature. *Brain* 130 (12): 3075–3090.

Pears, D. (1984). *Motivated Irrationality*. Oxford University Press.

Prigatano, G.P., ed., (2010). *The Study of Anosognosia*. Oxford University Press.

Ramachandran, V.S. (1995). Anosognosia in parietal lobe syndrome. *Consciousness and Cognition* 4(1): 22–51.

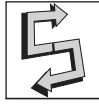
Ramachandran, V.S. (1996). The evolutionary biology of self-deception, laughter, dreaming and depression: Some clues from anosognosia. *Medical Hypotheses* 47(5): 347–362.

Ramachandran, V.S. and Blakeslee, S. (1998). *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. Morrow & Co.

Sartre, J.-P. (1949/1957). *Being and Nothingness*, trans. H. Barnes. Pocket Books.

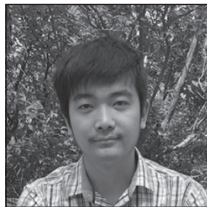
Unsworth, C. A. (2007). Cognitive and Perceptual Dysfunction. In T. J. Schmitz & S. B. O'Sullivan, eds., *Physical Rehabilitation*. Davis Company.

Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. Oxford University Press.



# P HENOMENAL CHARACTER, REPRESENTATIONAL CONTENT, AND THE INTERNAL CORRELATION OF EXPERIENCE: ARGUMENTS AGAINST TRACKING REPRESENTATIONALISM

**Bin Zhao** – Department  
of Philosophy,  
Shandong University  
E-mail:  
zhaobin529969263@  
gmail.com



Tracking representationalism is the theory that phenomenal consciousness is a matter of tracking physical properties in an appropriate way. This theory holds that phenomenal character can be explained in terms of representational content, and it also entails that there is unlikely to be a strong correlation between phenomenal character and neural states. However, the empirical evidence shows that both claims cannot be true. So, tracking representationalism is wrong. Its fault is due to ignoring the internal correlation of experience, the existence of which shows that phenomenal character is shaped by neural states to a large extent, so it cannot be wholly explained by representational content.

**Keywords:** tracking representationalism; phenomenal character; representational content; the internal correlation of experience.

## I. Introduction

Tracking representationalism is a contemporary theory of the phenomenal character of experience. To be specific, this theory holds that experiences are states representing the world in an appropriate way, and their phenomenal character can be explained by the representational contents. In addition, how experiences obtain their representational contents can be explained in terms of a tracking relation to physical properties [Bourget and Mendelovici, 2014: 210]. If this theory is correct, then we are able to put forward a physicalist account of phenomenal character and provide a simple answer to the “hard problem” of explaining why there is phenomenal character for a subject in experience [Chalmers, 1995: 201–202]. That is why it attracted a great deal of proponents, including Fred Dretske, William Lycan, and Michael Tye, in the recent years.

In this paper, I argue against tracking representationalism by using empirical evidence as well as a priori methods. My paper assumes the following structure. In section II, I sketch tracking representationalism and the motivation for it – the naturalization of phenomenal character. In section III, I put forward my first argument, according to which there is little correlation between phenomenal character and representational content, so the core thesis of tracking representationalism that phenomenal character can be explained in terms of representational content is untenable. In section IV, I put forward my second argument, according to which the internal correlation (viz., the isomorphic relationship between the structure of phenomenal



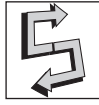
character and the structure of neural states) does exist, but that conflicts with the implication of tracking representationalism that there is unlikely to be such a correlation between phenomenal character and neural states. Finally, in section V, I elaborate that tracking representationalists' failure is due to their ignorance of the internal correlation. The existence of the internal correlation shows that phenomenal character is shaped by neural states to a large extent, so it cannot be wholly explained in terms of representational content.

## II. Tracking Representationalism and the Naturalization of Phenomenal Character

Tracking representationalism is a theory of phenomenal consciousness, the “what it is like” of experience. It aims to understand consciousness in terms of intentionality. Roughly speaking, intentionality is the aboutness or directedness of mental states. For example, we can believe that water is H<sub>2</sub>O rather than XYZ, or desire that it will rain tomorrow. These mental states both exhibit a kind of aboutness or directedness to the world. What a mental state is about or directed to is its representational content.

At a minimum, representationalism holds that experiences with the same representational content have the same phenomenal character and experiences with different phenomenal character have different representational contents. That is to say, phenomenal character supervenes on representational content. This fits well with our common sense. Suppose you are looking at a red rose, your experience of this rose has a particular phenomenal character. For the sake of simplicity, I will call this phenomenal character “reddish phenomenal character”. And your experience also has a particular representational content. If the color of the rose changes and the phenomenal character of your experience changes with it, then your experience will also have a different representational content [Bourget and Mendelovici, 2014: 219]. Nevertheless, why is the difference in phenomenal character accompanied by a difference in representational content? A rough explanation of this observation is that phenomenal character is a specific representational content or representational content meets certain further conditions [Tye, 2002: 137]. For example, in the above case, your experience has a reddish phenomenal character in virtue of the fact that it represents the sensible quality red, in other words, in virtue of the fact that the sensible quality red enters into the representational content of your experience.<sup>1</sup>

<sup>1</sup> Besides sensible qualities, the shape, size, location, and spatial relationships of external objects also enter into the representational content. But that has little relevance to our discussion.



Since experiencing a reddish phenomenal character is just representing the sensible quality red, what are sensible qualities? As a reductive version of representationalism, tracking representationalism advocates a physicalist account of sensible qualities, according to which sensible qualities are response-independent physical properties. For example, colors are surface reflectance, smell qualities and taste qualities are chemical properties of odorants and foods, auditory qualities are constituted by physical properties involving frequency, amplitude, and wave shape, and pain is tissue damage [Pautz, 2014a: 239]. In short, tracking representationalism holds that sensible qualities which constitute representational content are physical properties. And because phenomenal character can be explained by the representational content. Thus experiencing a specific phenomenal character is just representing a specific physical property or a complex of physical properties.

In addition, tracking representationalism advocates the tracking theory of intentionality, according to which, “intentionality is a matter of . . . correlating with features of the environment” [Bourget and Mendelovici, 2014: 215]. In detail, as Cutter and Tye illustrate:

[A] state represents a property or state of affairs in virtue of causally co-varying with that property or state of affairs under typical or optimal conditions. . . . Tokens of a state *S* in an individual *x* represent that *p* in virtue of the fact that: under optimal conditions, *x* tokens *S* iff *p*, and because *p* [Cutter and Tye, 2011: 90–91, emphasis in original].

That is to say, representation relation (the relation between representational vehicle and the property or state of affairs being represented) can be understood in functionalist terms. Therefore, the fact that experience represents physical properties is demystified.

In short, tracking representationalism is the combination of the following two doctrines:

Representationalism: The phenomenal character of an experience (at a minimum) supervenes on its representational content, and the former can be explained by the latter.

Tracking theory of intentionality: A state represents a property or state of affairs in virtue of causally co-varying with that property or state of affairs under optimal conditions [Cutter and Tye, 2011: 90].

If this theory is correct, then we can provide an easy answer to the hard problem. The hard problem is the problem of explaining why an experience should have the specific phenomenal character it does, rather than having some other phenomenal character or having no phenomenal character at all. For example, when you see something red, you usually experience a reddish phenomenal character. In virtue of what does your experience have the reddish phenomenal character rather than having a throbbing painfulness or having no phenomenal character at all? Chalmers argues that functional explanation does not apply to the hard problem, because “it is not a problem about the performance of functions. The problem persists



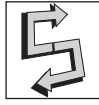
even when the performance of all the relevant functions is explained” [Chalmers, 1995: 201]. Moreover, we ordinarily suppose that perceptual experience is in the brain. But in the brain, we can only find grey matter and white matter, and there is no place for us to find phenomenal character. Where is phenomenal character located? Tracking representationalism provides a promising answer to these problems. According to this theory, phenomenal character is not an intrinsic property of experience, so we cannot find it in the brain. In fact, it is a representational property of experience. It can be explained in terms of the representational content which is constituted by physical properties represented by the experience. Moreover, on tracking representationalism, representation relation can be explained in functionalist terms. Therefore, the fact that there is “something it is like” for a subject in experience is no longer a myth.

In brief, tracking representationalism is a two-step strategy of the naturalization of mind. The first step – representationalism – accounts for phenomenal character in terms of representational content. The second step – tracking theory of intentionality – accounts for representational content in terms of a tracking relation to the physical properties [Bourget and Mendelovici, 2014: 210]. Therefore, phenomenal character is naturalized and the hard problem is solved.

### III. Phenomenal Character and Representational Content

According to tracking representationalism, the phenomenal character of an experience can be explained by its representational content. This radical idea aroused considerable controversy. Philosophers constructed many thought experiments to refute it such as “the brain in vat”, “swampman”, “inverted spectrum”, and “inverted earth”.<sup>2</sup> What they have in common is denying that phenomenal character supervenes on representational content. Thus, denying that the former can be explained by the latter. For example, in the thought experiment “inverted spectrum”, visual experiences of two people can represent the same surface reflectance, while one has a reddish phenomenal character, the other has a greenish phenomenal character [Shoemaker, 1994: 34]. So phenomenal character does not supervene on representational content and, therefore, the former cannot be explained by the latter. In tracking representationalists’ view, although scenarios in these thought experiments are conceivable, they are metaphysically impossible. It is just like the case where we can conceive gold is not element 79, but that is impossible [Tye, 1986: 5–6]. In brief, the intuition these

<sup>2</sup> See: Horgan and Tienson (2002), Block (1998), Shoemaker (1994), and Block (1990) for related discussions.



thought experiments rely on is controversial, so conclusions drawn from them are unreliable.

Unlike these thought experiments, I will argue against tracking representationalism by using empirical evidence in this section. Contrary to tracking representationalism, empirical evidence shows that there is little correlation between phenomenal character and representational content, so phenomenal character cannot be explained in terms of representational content. Thus empirical evidence poses a challenge to tracking representationalism. I will use olfactory experience as an example to illustrate this challenge in detail.<sup>3</sup>

When we smell a smell, we usually experience some phenomenal character. On tracking representationalism, these phenomenal character is determined by and can be explained by the representational content of experience. But research shows that there is little correlation between phenomenal character and representational content, so we are not able to explain phenomenal character in terms of representational content. For example, although hydrogen sulfide (molecular formula:  $H_2S$ ) and borane (molecular formula:  $BH_3$ ) both smell like rotten eggs, they have totally different molecular structures, in this case, two experiences have very similar phenomenal character, but their representational contents are totally different;<sup>4</sup> although L-carvone and D-carvone have very similar molecular structures (they are stereoisomers – molecules with exactly the same molecular formula and slightly different structures, just like left hand and right hand), they smell very different – the former smells like spearmint, while the latter smells like caraway [Da Costa and Eri, 2009: 27], in this case, two experiences have totally different phenomenal character, but their representational contents are very similar. Since phenomenal character cannot be explained by representational content, tracking representationalism is refuted directly.

To this refutation, tracking representationalists have several possible replies. The first reply says what is represented by the olfactory experience is some chemical property other than the structure of odor molecule. The second reply says that every odor is a disjunction of some odor molecules rather than a specific structure of odor molecule, and it is this disjunction that is represented by the olfactory experience [Lycan, 2014: 7]. In what follows, I will argue these two replies cannot save tracking representationalism in the face of my challenge.

<sup>3</sup> For the sake of simplicity, I only focus on olfactory experience in this section, though the thesis that phenomenal character cannot be explained in terms of representational content applies to experiences of all sensory modalities such as vision, audition, and gustation. See Pautz (2014a) and Goldstein (2010) for further discussions.

<sup>4</sup> Though many tracking representationalists maintain that olfactory experience represents qualities of odorants, for example, Tye (2002), as far as I know, none of them have developed a detailed view (Batty (2009) may be an exception. Nonetheless, I do not know whether she is a tracking representationalist, though she has developed a representational account of olfactory experience). For the purpose of discussion, I assume that olfactory experience represents the structure of odor molecule in this passage. This may be the most promising view on the side of tracking representationalists.



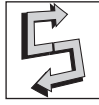
Now let us turn to the first reply, according to which what is represented by the olfactory experience is some chemical property other than the structure of odor molecule. So which kind of chemical property is the “suitable candidate”? Tracking representationalists may turn to the vibrational theory of olfaction to defend their theory. According to the vibrational theory, what is perceived by the olfactory experience is the vibration frequency of chemical bond of odor molecule. This theory can explain the fact that hydrogen sulfide and borane have the same odor though their molecular structures are totally different, and this is because the S-H bond of hydrogen sulfide and the B-H bond of borane have very close vibration frequencies [Turin, 2009: 261]. So olfactory experiences of these two odor molecules have very similar representational contents, no wonder they have very similar phenomenal character.

Nevertheless, the explanatory power of the vibrational theory of olfaction is very limited, as there are many facts which cannot be explained by it. For instance, it cannot explain the case of carvone mentioned above. L-carvone and D-carvone not only have very similar molecular structures, but also their corresponding chemical bonds have exactly the same vibration frequencies, however they smell different. In this case, two olfactory experiences have totally different phenomenal character, but their representational contents are the same.

Moreover, in some cases, the change in the concentration of an odor molecule can alter the phenomenal character of olfactory experience. For example, “the odor of thioterpineol is described as ‘tropical fruit’ at a low concentration, as ‘grapefruit’ at a higher concentration, and as ‘stench’ at a still higher concentration” [Pautz, 2014a: 245]. However, concentration cannot alter the vibration frequencies of chemical bonds. In this case, we have olfactory experiences which have different phenomenal character and the same representational content. Therefore, phenomenal character cannot be explained by representational content.

According to the second reply, every odor is a disjunction of some odor molecules rather than a specific structure of odor molecule. For example, the odor “smells like rotten eggs” can be characterized as “hydrogen sulfide v borane v sulfur monoxide v ... (this disjunction may be infinite)”, just as some tracking representationalists hold the idea that color is a disjunctive microstructural property of objects [Lycan, 1996: 73]. In addition, it is this disjunction that is represented by the olfactory experience. So in the case of hydrogen sulfide and borane, two olfactory experiences represent the same disjunction, thus they have the same representational content. That’s why they also have the same phenomenal character.

However, this reply cannot eliminate all the counterexamples. For instance, “androstadienone . . . is perceived by different individuals as offensive (‘sweaty, urinous’), pleasant (‘sweet, floral’) or odourless” [Keller et al., 2007: 468]. If the case is “androstadienone is perceived by different individuals as offensive or odourless”, then tracking representationalists could



explain it. Their explanation will be that the first group of people represented the disjunction “androstadienone v ...”, so they smelled something offensive; while the latter group of people failed to represent this disjunction due to some dysfunction, so they did not smell anything.

Nonetheless, this explanation does not work for the case here, because there is another group of people who smelled something pleasant. What is represented by them? Firstly, we have no reason to assume that these people were in states of misrepresentation, viz., representing the odor molecule as having a property which it does not possess, because the test is conducted under normal conditions and the subjects do not have physiological or cognitive defects. Secondly, we have no reason to suppose that they represented other odor molecules or another disjunction of odor molecules because only pure androstadienone is used in this test. Therefore, even resorting to the disjunction of odor molecules cannot save tracking representationalism from my argument.

The above analysis shows that there is little correlation between phenomenal character and representational content. Thus, the core thesis of tracking representationalism that phenomenal character can be explained in terms of representational content is untenable, even resorting to ad hoc hypotheses like “every odor is a disjunction of some odor molecules” cannot defend it.

#### IV. The Internal Correlation of Experience

In section II, I have mentioned that tracking representationalism advocates the tracking theory of intentionality, according to which a state represents another state in virtue of the fact that there is a causal covariant relation between them rather than they have some properties in common. For instance, we can hardly say that the word “cup” and a cup have anything in common, but the former can still represent the latter. On tracking representationalism, this idea also applies to experiences such as pain. As the representational vehicle, the neural correlate of a pain experience represents a tissue damage of a specific intensity in virtue of the fact that the former is caused by the latter under optimal conditions rather than they have some properties in common.<sup>5</sup> Even if the neural state did not have the properties which it actually has, for example, if it had a different firing rate, it can still represent the tissue damage of a specific intensity as long as there is an appropriate causal covariant relation between it and the tissue damage.

Therefore, the firing rates of neural states are unlikely in direct proportion to the intensities of tissue damage. In addition, tracking representatio-

<sup>5</sup> In this paper, by “pain” I mean the sensory component of pain rather than the affective component of pain. For related discussions, see Goldstein (2010).



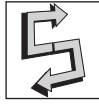


nalism holds that pain intensity is determined by the intensity of tissue damage represented by the pain experience, so pain intensities are unlikely in direct proportion to the firing rates of neural states [Pautz, 2014b: 308]. That is to say, there is unlikely to be an isomorphic relationship between the structure of phenomenal character (viz., the similarities or numerical relationships between phenomenal character) of pain experience and the structure of neural states (viz., the similarities or numerical relationships between neural states). For the sake of simplicity, call this isomorphic relationship “the internal correlation” [Pautz, 2014a: 242]. In sum, tracking representationalism implies that the internal correlation is not possible to exist.

However, the above implication conflicts with neuroscientists’ discovery that pain intensity closely matches firing rates of neurons in the primary somatosensory cortex [Timmermann et al., 2001: 1501]. So the internal correlation does exist. In addition, neuroscience research shows that this thesis applies to experiences of all sensory modalities such as vision, olfaction, gustation, and audition [Pautz, 2014a; Goldstein, 2010]. Thus, contrary to the implication of tracking representationalism, there is a strong correlation between the phenomenal character of experience and neural states.

A possible response of tracking representationalists is that the internal correlation is just a coincidence because it does not exist in some counterfactual cases. Its reason is as follows: What an experience normally represents is what nature designed it to represent, what it has as its biological purpose to represent [Tye, 1998: 461]. For instance, we have reddish phenomenal character in some cases because some of our brain states have the biological function of representing red objects. And what biological function we have depends on our evolutionary history [Dretske, 1995: 7]. If our evolutionary history was slightly different, these brain states could be designed to represent blue objects rather than red objects. Thus, even if two individuals have the same internal states, their internal states can still represent different physical properties if the species they belong to have different evolutionary histories [Tye, 2009: 195].<sup>6</sup> Likewise, neural states of the same firing rate can represent tissue damages of different pain intensities if their subjects have different evolutionary histories. In the real case, the nature accidentally designed a neural state whose firing rate is  $n$  units to represent tissue damage whose intensity is  $kn$  units ( $k$  is a constant). This leads to the fact that pain intensities are in direct proportion to the firing rates of neural states. While in some counterfactual cases, the nature designed a neural state whose firing rate is  $n$  units to represent tissue damage whose intensity is not  $kn$  units. Then pain intensities are not in direct proportion to the firing rates of neural states. So pain intensities are in direct proportion

<sup>6</sup> Here is a simple analogy: the same voltmeter can perform different functions under different designs. It can be designed as a speedometer, then to represent the speed of a car; it also can be designed as a fuel gauge, then to represent the volume of oil. In these two cases, the internal states of the voltmeters are the same, but what they represent are different. For detailed discussions; see: Tye (2009).



to the firing rates of neural states in some cases is not attributable to the fact that there is a strong correlation between them. Their ostensible strong correlation is accidentally caused by the design on the earth.

The above response tries to explain away the correlation between phenomenal character and neural states by resorting to evolutionary history. I will use a thought experiment to argue that this response implies that phenomenal character has no causal effects on the physical. But there is some tension between this implication and some tracking representationalists' claim that "phenomenal character... [can] make a direct impact on the beliefs/desires system" [Tye, 1995: 138].<sup>7</sup> Thus the above response fails as a suitable response of tracking representationalists. My thought experiment is as follows:<sup>8</sup>

Oscar is an ordinary man, and Twin Oscar is his microphysical duplicate who was created in a chemical reaction caused by a lightning. At time  $t$ , they were respectively transported to planet  $A$  and planet  $B$  which are suitable for human habitation. Suppose there are the same physical events on planet  $A$  and planet  $B$  at time  $t$ , and these two planets obey the same physical laws. If causal closure applies to the physical world, then in the following time, planet  $A$  and planet  $B$  will be identical in all physical aspects.<sup>9</sup> As an ordinary man, Oscar can experience phenomenal character. While Twin Oscar does not have an evolutionary history and nature did not make any design on him. So his internal states cannot play the role of representation. As a consequence, he cannot experience phenomenal character. Thus there are the same physical events and different phenomenal character on planet  $A$  and planet  $B$  after time  $t$ . The best explanation to this conclusion is that phenomenal character has no causal effects on the physical.

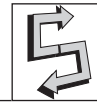
This implication totally violates our common sense. And, as I mentioned above, it is also in tension with some tracking representationalists' claim that phenomenal character has a direct impact on beliefs (or desires) [Tye, 1995: 138].<sup>10</sup> (For example, when we experience a reddish phenomenal character under normal conditions, we usually form a belief that we have seen something red. So the reddish phenomenal character (or the experience which has a reddish phenomenal character) has causal effects on our beliefs.) Thus the above response, which aims at explaining away the

<sup>7</sup> This claim aims to demarcate between representation with phenomenal character (for example, experience) and representation without phenomenal character (for example, desire and belief). See: Tye (1995) for related discussions.

<sup>8</sup> This thought experiment derives from the "swampman" thought experiment in Davidson (1987).

<sup>9</sup> The influence external environment has on these two planets is ignored, as it has little relevance to the discussion here.

<sup>10</sup> There is a possible solution to dissolve the tension. This solution maintains that though phenomenal character (or the experience which has a phenomenal character) has causal effects on beliefs (or desires), while beliefs and desires do not have causal effects on the physical, so phenomenal character (or the experience which has a phenomenal character) does not have (direct or indirect) causal effects on the physical. However, this solution also violates our common sense. As far as I know, tracking representationalists have not advocated such a strange view on belief and desire.



internal correlation, fails as a suitable response of tracking representationalists. This means that tracking representationalists are unable to explain away the internal correlation which is incompatible with their theory. So tracking representationalism is in trouble.

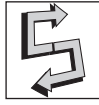
## V. Where Tracking Representationalism Went Wrong

In the last two sections, I argued that (a) phenomenal character cannot be explained in terms of representational content, and (b) there is a strong correlation between phenomenal character and neural states, which conflicts with the implication of tracking representationalism, so tracking representationalism fails. In my view, its failure is due to ignoring the internal correlation mentioned above, the existence of which shows that phenomenal character is shaped by neural states to a large extent, so it cannot be wholly explained in terms of representational content.

Take color experience for instance, opponent process theory of color vision holds that the phenomenal character of color experience can be explained by activation levels of opponent processing channels. Opponent processing channels consist of red-green opponent channel and yellow-blue opponent channel. When red-green opponent channel assumes a positive state of activation, we have a reddish experience; when it assumes a negative state of activation, we have a greenish experience. When yellow-blue opponent channel assumes a positive state of activation we have a yellowish experience; when it assumes a negative state of activation we have a bluish experience. Thus, according to opponent process theory, the phenomenal character of color experience is determined by activation levels of red-green opponent channel and yellow-blue opponent channel. Though just resorting to representation of surface reflectance cannot explain why there does not exist a binary color combined of red and green or a binary color combined of yellow and blue, opponent process theory can explain why these binary colors do not exist. This is because neither of the two channels can be in a state of both activation and inhibition at the same time [Pautz, 2006].

Thus it can be seen that though our color experience is a representation of physical properties (surface reflectance), its phenomenal character cannot be wholly explained by representational content. On the contrary, many features of phenomenal character (for instance, the structure of phenomenal character of color experiences) can only be explained by the internal states (for instance, the activation levels of opponent processing channels), which shows that internal states play an important role in determining the phenomenal character of experience [Pautz, 2006: 228].

In sum, experience is like footprints in the snow. The formation of footprints is not only related to the texture of snow, but also related to the shape



of foot; which phenomenal character an experience has is not only related to its representational content, but also related to the internal states of the individual. In other words, the phenomenal character of experience is the product of the interaction between internal factors and external factors. Tracking representationalism duly appreciates the representational character of experience, but not the internal correlation of experience. This leads to the mistaken ideas that the function of neural states is just representation and phenomenal character can be explained by representational content. However, in fact, neural states shape our phenomenal character to a large extent, so the relationship between them cannot be the relationship between representational vehicle and representational content. Thus, we are not able to explain phenomenal character in terms of representational content.

## VI. Conclusion

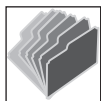
Tracking representationalism aims to put forward a naturalistic account of phenomenal character. If this theory is correct, then phenomenal character is determined by and can be explained by representational content which in turn can be explained in terms of a tracking relation to physical properties. Given my arguments, tracking representationalism is confronted with insurmountable problems. It cannot defend its core thesis that the phenomenal character of an experience can be explained by its representational content, and it also conflicts with neuroscientists's discovery that there is a strong correlation between phenomenal character and neural states (viz., the internal correlation). So tracking representationalism fails as a suitable naturalistic theory of mind. Its fault is due to ignoring the internal correlation of experience, the existence of which shows that phenomenal character is shaped by neural states to a large extent, so it cannot be wholly explained by representational content. Thus we should give up tracking representationalism and seek a theory which is compatible with the internal correlation.

## References

- Batty, Clare. 2009. "What's That Smell?" *Southern Journal of Philosophy*, 47 (4): 321-348.
- Block, Ned. 1990. "Inverted Earth." *Philosophical Perspectives*, 4: 53-79.
- . 1998. "Is Experience Just Representing?" *Philosophy and Phenomenological Research*, 58 (3): 663-670.
- Bourget, David and Angela Mendelovici. 2014. "Tracking Representationalism: William Lycan, Fred Dretske, and Michael Tye." In Andrew Bailey (ed.), *Philosophy of Mind: The Key Thinkers*. London: Bloomsbury Publishing.



- Chalmers, David J. 1995. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies*, 2 (3): 200-219.
- Cutter, Brian and Michael Tye. 2011. "Tracking Representationalism and the Painfulness of Pain." *Philosophical Issues*, 21 (1): 90-109.
- Da Costa, Neil C. and Sanja Eri. 2009. "Identification of Aroma Chemicals." In David J. Rowe (ed.), *Chemistry and Technology of Flavors and Fragrances*. Oxford: Blackwell Publishing.
- Davidson, Donald. 1987. "Knowing One's Own Mind." *Proceedings of the American Philosophical Association*, 60 (3): 441-458.
- Dretske, Fred. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Goldstein, E. Bruce. 2010. *Sensation and Perception*. Belmont, CA: Wadsworth.
- Horgan, Terence and John Tienson. 2002. "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In David J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.
- Keller, Andreas, Hanyi Zhuang, Qiuyi Chi, Leslie B. Vosshall, and Hiroaki Matsunami. 2007. "Genetic Variation in a Human Odorant Receptor Alters Odour Perception." *Nature*, 449 (7161): 468-472.
- Lycan, William. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- . 2014. "The Intentionality of Smell." *Frontiers in Psychology*, 5: 1-8.
- Pautz, Adam. 2006. "Sensory Awareness is not a Wide Physical Relation: An Empirical Argument Against Externalist Intentionalism." *Nous*, 40 (2): 205-240.
- . 2014a. "The Real Trouble with Phenomenal Externalism: New Empirical Evidence for a Brain-Based Theory of Consciousness." In Richard Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. Dordrecht: Springer Netherlands.
- . 2014b. "Ignoring the Real Problems for Phenomenal Externalism: A Reply to Hilbert and Klein." In Richard Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. Dordrecht: Springer Netherlands.
- Shoemaker, Sydney. 1994. "Phenomenal Character." *Nous*, 28: 21-38.
- Timmermann, Lars, Markus Ploner, Katrin Haucke, Frank Schmitz, Rudiger Baltissen, and Alfons Schnitzle. 2001. "Differential Coding of Pain Intensity in the Human Primary and Secondary Somatosensory Cortex." *Journal of Neurophysiology*, 86: 1499-1503.
- Turin, Luca. 2009. "Rational Odorant Design." In David J. Rowe (ed.), *Chemistry and Technology of Flavors and Fragrances*. Oxford: Blackwell Publishing.
- Tye, Michael. 1986. "The Subjective Qualities of Experience." *Mind*, 95 (377), 1-17.
- . 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
- . 1998. "Inverted Earth, Swampman, and Representationalism." *Philosophical Perspectives*, 32:459-78.
- . 2002. "Representationalism and the Transparency of Experience." *Nous*, 36 (1): 137-151.
- . 2009. *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge, MA: MIT Press.



## **А**ЛЕКСАНДР БЭН И ЕГО ФИЛОСОФСКИЕ ВОЗЗРЕНИЯ. ПРЕДИСЛОВИЕ К ПЕРЕВОДУ<sup>1</sup>

**Андрей Анатольевич Веретенников** –  
кандидат философских наук, научный со-  
трудник сектора современной западной  
философии Института философии  
РАН. Email: and-veret@gmail.com

**Andrey Veretennikov** – Institute  
of Philosophy RAS.

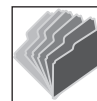
## **А**LEXANDER BAIN AND HIS PHILOSOPHICAL VIEWS

Один из самых известных мыслителей второй половины XIX в., Александр Бэн (1818–1903) является одним из столпов ассоциативной психологии – доминировавшего в то время метода в описании психических явлений. Наряду с В. Вундтом (1832–1920) он известен как лидер движения за отделение психологии как науки, «экспериментальной психологии», от философии. В то же время его деятельность выходила далеко за пределы того, что мы сегодня могли бы назвать собственно исследованиями по психологии. Начав свою карьеру как автор научно-популярных произведений и учебников по самым разнообразным предметам – от астрономии до английской грамматики, он занимает кафедру физики в Университете Глазго (1845–1848), а с 1860 г. – кафедру логики в только что созданном Абердинском университете. В 1881–1887 гг. он становится ректором этого учебного заведения, выведя его в ряд ведущих вузов Шотландии и всего Соединенного Королевства.

Деятельность Бэна в научной, образовательной и просветительской областях нашла отражение в его работах по теории образования. Знакомство с трудами основателя эволюционной теории Ч. Дарвина (1809–1882) и отца философского эволюционизма Г. Спенсера (1820–1903), перенесшего методы изучения естественного отбора в область психического, связав его тем самым с окружающей средой, также повлияло на мировоззрение Бэна.

Эволюционный принцип в объяснении генезиса психических явлений изменил представления о развитии человеческой психики в сторону натурализма Д. Юма и И. Бенета многих современников Бэна, но последний является одним из его наиболее убежденных сторонников. Будучи другом и соратником Дж. Ст. Милля (1806–1873), он во многом перенял и модифицировал методы индуктивной логики из знаменитого труда философа «Система логики силлогистической и индуктивной» (1843). У. Гамильтон (1788–1856), выдающийся представитель шотландской философии, сочетавший в своих работах идеи Т. Рида и И. Канта, был для Бэна одновременно и объектом критики,

<sup>1</sup> Подготовлено при поддержке РФНФ, проект № 14–33–01043.



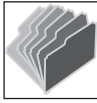
и представителем того типа «трезвого» мышления, который характеризует шотландскую школу философии здравого смысла. В частности, это сказалось на представлениях Бэна о непосредственном, или «безусловном», в терминологии Гамильтона, само существование которого опровергает, с точки зрения ряда мыслителей, классические построения философского идеализма. Близость исследований в области логики XIX в., понимаемой во многом как теория познания, и психологии, постепенно переходящей от описательных, в том числе интроспективных, методов к методам экспериментальным, отразилась и на работах Бэна. В последней четверти XIX в. Бэн становится одним из самых известных психологов, одинаково популярных и среди читающей публики, и среди ученых всего мира. Стоит указать также и на влияние мыслителя в России XIX в.: многие из его работ, включая учебники и выдержки из различных произведений, переводились и печатались в течение буквально нескольких лет после издания на языке оригинала, что показывает их значение не только для научного сообщества, но и для содержания курсов по психологии в России<sup>2</sup>.

Для понимания воззрений Бэна на природу психофизической связи нужно указать на специфический метод познания, который психолог использовал для описания динамики и соотношения языка, психических и физических явлений. Этим методом он во многом обязан Миллю, фундаментальная «Система логики» которого изменила представления не только о методах познания в области «моральных», или социальных, наук, но и в науке о человеческой душе. Разделение Миллем психологии и философии по методу, а не по предмету оказало влияние на последующую психологию, в том числе и на поиск Бэном нового метода в области исследования психики. Милль ввел причинность в методы индукции, например в пятом правиле «сопутствующих изменений» (*concomitant variations*): «Если за изменением одного явления наблюдается изменение другого, то мы можем заключить о причинной связи между ними», что модифицируется Бэном в описание психофизической связи.

Каждому психическому факту сопутствует (*concomitance*) факт физический, что показывает их неразрывную связь и опровергает выводы «философов» о независимом существовании духовных явлений, не сопровождаемых физическими<sup>3</sup>. Это положение ставит его в ряд с

<sup>2</sup> Среди них можно назвать: *Об изучении характера*. СПб., 1866; *Стилистика и теория устной и письменной речи*. М., 1866; *Психофизиологические этюды*. СПб., 1869; *Душа и тело*. СПб., 1881; *Психология*. СПб., 1881; 1887; *Наука о воспитании*. СПб., 1881. В постсоветскую эпоху работа Бэна «Психология», содержащая выдержки из прочих его трудов по психологии, была переиздана в книге: *Основные направления психологии в классических трудах. Ассоциативная психология*. Г. Эббингауз. *Очерк психологии*. А. Бэн. Психология. М.: АСТ, 1998. 544 с. Дальнейшие цитаты даются по этому изданию.

<sup>3</sup> «Положение, что физические процессы неизменно сопутствуют психическим, есть закон всего духа». *Бэн А.* Психология. М., 1998. С. 215.



другими более ранними сторонниками психофизического параллелизма и некоторыми современными философами сознания, использующими понятие сопровождения (supervenience) для описания связи между психическими и физическими свойствами. Стандартной проблемой для такого подхода в наши дни является проблема эпифеноменальных качеств, заключающаяся в том, что если психическое всегда сопровождается физическим, а физическое выполняет всю известную нам каузальную работу, то на долю психического не остается ничего, оно будет каузально инертным. Бэн решает эту проблему при помощи описания причинности как сопутствия, а не необходимой связи. Это, учитывая все соответствующие такому пониманию проблемы вероятности связи, позволяет обойти некоторые трудности, возникающие в области не только тезиса о психофизическом параллелизме, но и в современной теории сопровождения<sup>4</sup>.

Однако определение самого духа или ментального у Бэна не является вполне однозначным. В зрелый период мыслитель отождествлял определение и классификацию, утверждая, что «определение духа есть вместе с тем и его разделение, т.е. классификация его главных и основных свойств»<sup>5</sup>. Среди этих свойств он называл «чувство и хотение» в их связи с ощущениями и способностью различения, сам ум, эмоции и волю, параллелизм процессов в духе и теле. В то же время он указывал на то, что слово «сознание», если оно обозначает все, что чувствуется, не может иметь никакого определения: «Подобно слову “бытие”, это искусственно выдуманный термин для объединения двух высших родов существующего – объекта и субъекта»<sup>6</sup>. Таким образом, «сознание» представляет для него в этот период наиболее общий, или бессодержательный, неанализируемый далее термин. Публикуемый перевод дает представление об изменении его взглядов в поздний период научной деятельности, а именно, о попытке дать недвусмысленное определение термина «сознание».

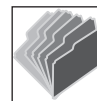
Помимо уже ставшего стандартным на тот период различения между центром осознания и его периферией – рефлексам и привычкой – в статье «Определение сознания» заметны как следы его разработок в области логики в духе Милля, так и нечто новое, а именно: попытка работать с терминами в новом для психологии и философии конца XIX в. ключе. Речь идет о проведении различия между терминологическим языком науки и обычным, разговорным языком. Тако-

<sup>4</sup> «Сложность духовных явлений не позволяет нам устанавливать для них законы безусловно всеобщие; поэтому во многих случаях нам приходится довольствоваться лишь более или менее вероятными обобщениями». «Это не основные законы природы, вроде закона тяготения или закона сохранения энергии; однако они вполне доказаны для всех проявлений духа и потому вполне приложимы в этих пределах». Указ. соч. С. 231 и 229.

<sup>5</sup> Там же. С. 212.

<sup>6</sup> Там же. С. 219.

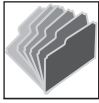




го рода подход широко использовался еще Ридом и Гамильтоном, но в поздних работах Бэна он приобретает новую окраску. Бэн берет за основу семантику обычного языка, или языка «народной психологии», которым люди обычно описывают свои психические состояния. По его мнению, значения терминов в таком использовании достаточно ясны и недвусмысленны, а попытка затуманить их (или ввести дополнительные коннотации в неконнотативные имена, в терминах Милля) вредна для психологии как науки.

В этом контексте интересно жесткое различие, проводимое Бэном между «научной» психологией и философией, доминировавшей в то время в университетах англоязычного мира, – философией абсолютного идеализма, представленной в Великобритании такими выдающимися мыслителями, как Ф. Брэдли (1846–1924) и его младший современник Дж. МакТаггарт (1866–1925), а в США – Дж. Ройсом (1855–1916). Эти философы, иногда также называемые английскими (американскими) неогегельянцами (например, Брэдли), отрицая эмпиризм, разделяли сферу познания на явления, в области которых возможно противоречие, и сферу реальности, или духа, где только и возможна абсолютная истина. Таким образом, любые исследования психического в духе британского эмпиризма, отталкивающиеся от наблюдения за областью осознания ощущений, оказывались исследованиями «второго сорта», результаты которых не могут претендовать на истинность в полном смысле слова. Бэн жестко отграничивает «научную» психологию от философии такого типа. Так как она в то время являлась основным типом философии, Бэн получает дополнительный аргумент в пользу разделения «науки психологии» и философии.

Заключительная часть публикуемой статьи, содержащая анализ термина «самосознание», дает представление о его взглядах и позволяет сравнить их с взглядами нового поколения философов, в частности Б. Рассела и Дж.Э. Мура. Метод анализа последнего во многом повторяет подход со стороны обыденного языка Бэна, но, пожалуй, главным здесь будет то, что у зачинателя «восстания против идеализма» был такой влиятельный предшественник, а значит, этот «переворот в философии» объясняется не только внутренними причинами в самой философии, но и научной психологией того времени.



## О ПРЕДЕЛЕНИЕ И ПРОБЛЕМЫ СОЗНАНИЯ<sup>1</sup>

Александр Бэн

## D EFINITION AND PROBLEMS OF CONSCIOUSNESS<sup>2</sup>

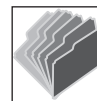
Alexander Bain

К настоящему времени можно считать, что процесс определения того, что есть сознание, вполне завершен. Сказанное вполне справедливо по отношению к теории, хотя на практике отклонения встречаются довольно часто. В подготовительной статье к своим «Опытам о способностях ума» Рид дает определение в соответствии с правилами, используемыми логиками; тем не менее Гамильтон обвиняет его в смешении *вербального* и *реального* определений. Это замечание Гамильтон делает в дополнение к своей ремарке о том, что индивидов нельзя определить (р. 220). «Как прекрасно отмечали логики древности, *Omnis intuitiva notitia est definitio*, т.е. *взгляд на саму вещь есть ее лучшее определение*. Это справедливо и для объектов чувственности, и для объектов самосознания». Кто именно из логиков древности является родоначальником этой формулы, я сказать не могу; мне неизвестны другие случаи использования этого высказывания. Понятое в строгом смысле, это выражение, как его использует Гамильтон, оказывается довольно ограниченным. Как может показаться, он понимает под ним актуальное наличие в восприятии вещи, которую нужно определить; однако эта интерпретация несовместима с его упоминанием объектов самосознания, так как последние, как известно, могут быть показаны не иначе как некоторым иносказательным, непрямым способом.

Общепринято, что определение неприменимо к исходным понятиям – ограничение, которого до сих пор не придерживаются наши создатели словарей. Дополнительное ограничение существует в области понятий, по своей природе составных или выводных, но которые невозможно выразить через их составные части. Таким образом, *жизнь, смерть, здоровье, болезнь и переживания* невозможно определить иначе, как только указав на конкретные примеры, известные тому, для кого дается определение. Если рассматривать определение в его более широком и неопределенном смысле передачи умопостижимых истин, выразимых в языке, фактически в качестве средства обыденного, а не научного объяснения, то его объем может быть определен посредством выявления согласия между различными людьми по отношению к объекту, обозна-

<sup>1</sup> Подготовлено при поддержке РГНФ, проект 14-33-01043.

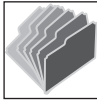
<sup>2</sup> Mind, New Series. Jul., 1894. Vol. 3. No. 11. P. 348–361.



чаемому данным именем. Если же даже в самых различных обстоятельствах люди в общем понимают одно и то же под тем, что подразумевается в процессе использования данного слова, то это слово определено для целей взаимопонимания и разъяснения любых сложных значений, в которые оно входит в качестве составной части. То, что существует множество подобных имен, явствует из возможности обращения с разумной речью к широким массам. Без сомнения, в области технических и глубоких материй используемые имена принадлежат к обычному, повседневному словарю, однако они употребляются с особыми ограничениями, которые уже должны быть известны обучаемым. В самом деле, в каждом разделе знания, которому придана научная форма, необходимо давать предварительные разъяснения, позволяющие избавить используемые имена от неопределенности, придаваемой им в обыденном словоупотреблении. Говорящий от имени науки с радостью использует подобные слова как вполне недвусмысленные сами по себе, т.е. понимаемые одинаково всеми, к кому он обращается. Такого рода слова будут вполне подходящим средством, чтобы выразить сложные и запутанные термины, которые в противном случае не могут быть поняты ясно и одновременно.

Следующие наблюдения более или менее применимы ко всему словарю ума, как он используется в *психологии*. Некоторые термины этого словаря понятны сами по себе и не требуют определения по единственной причине: они указывают на факты или явления, настолько знакомые и настолько недвусмысленные, что мы все согласны с их значением. Таким образом, они оказываются краеугольными при определении или объяснении прочих более многочисленных классов терминов, выражающих важные обобщения высокого уровня. Их значение достаточно запутано, и они требуют всех средств методичного определения. Приведем примеры двух этих типов после указания общего направления настоящей статьи, в которой речь пойдет о ведущем термине «сознание».

Это слово в самых разных случаях на удивление свободно от двусмысленности и вполне прозрачно для обыденного понимания. Однако с ним связано множество тонких и сложных проблем и в силу этого следствия из них не так просто распутать. Его сущность настолько обширна, что оно кажется включающим всю нашу психическую жизнь; оно является именно тем родовым словом, под которым собираются все разнообразные виды нашей психической активности. В силу этого мы вполне можем предположить, что с ним связаны все великие проблемы психологии. В то же время как бы оно ни было обширно, есть веские основания для того, чтобы не делать его основным термином всей психологии; последнее станет ясно в ходе нашего рассмотрения сферы его применения. Мы хотим показать, что существуют определенные проблемы, более связанные с этим, а не другими именами; в то



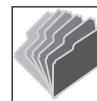
же время у нас будет возможность коснуться некоторых проблем, которые более тесно и непосредственно связаны с другими словами.

Позвольте мне кратко представить два класса терминов, на которые мы уже указывали как входящие в словарь ума. В первый класс слов, понимаемых всеми в одном и том же смысле, входит и должна упоминаться в первую очередь крайне важная пара – *удовольствие* и *боль*. Принимая за данность то, что мы достаточно наблюдательны, чтобы замечать происходящее в нашем интроспективном сознании и осознавать боль и удовольствие, иногда испытываемые нами, мы согласны с другими относительно этих фактов нашего опыта. Мы не путаем удовольствие ни с болью, ни с состоянием сознания. Свойства мира *объектов* со всей их отчетливостью редко когда бывают более ясными или несомненными, чем эти два основных свойства нашей психической жизни. Следовательно, используя эти термины, не требующие определений сами по себе, мы можем внести ясность в значение менее определенной терминологии ментального.

Действие, называемое *ощущением сходства, подобия, узнавания, чувства согласия*, также является прекрасно определенным фактом нашей психической природы, который мы не путаем ни с чем другим. Когда мы говорим, что две вещи одинаковы в нашем схватывании, мы указываем на подлинно ментальный акт, а слушающие нас понимают утверждение именно так, как мы хотели.

Другое имя, представляющее вполне ясный процесс, знакомый с раннего детства и разделяемый с другими, – это процесс, называемый *памятью, воспоминанием, удержанием в уме*. Этот подлинно психический процесс, одновременно типичный и отличающийся от всех других, содержит значительную часть нашей ментальной жизни. Язык предоставляет нам множество эквивалентов или синонимов для этой замечательной способности, большинство из которых вполне понятны и недвусмысленны. Среди них *обучение, забывание, приобретение, выучивание наизусть, учение, натаскивание* и т.д. Использование любого из этих имен указывает на привычный факт опыта; они не нуждаются в словарном определении и понимаются вполне ясно, так как указывают на знакомые каждому переживания.

На этом мы еще не закончили наше перечисление терминов, принадлежащих субъективному миру, являющихся настолько недвусмысленными, что сама возможность их непонимания исключается. В именах *действия, деятельности* и их противоположностях *бездействия* или *пассивности* мы также обнаруживаем основу для всеобщего согласия по поводу ментальных фактов. Действие, без сомнения, применимо и к силам материального мира, однако оно является термином из области ментального, никоим образом не затуманенным другими сферами его применения.



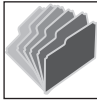
Также очевидна важность другой известной и отличающейся от всего прочего пары терминов, принадлежащей как психической, так и телесной жизни, – это *сон* и *пробуждение*. Касательно значения этих терминов расхождений быть не может.

Указав на некоторые из основных терминов базового словаря, необходимо также упомянуть в целях иллюстрации по контрасту некоторые примеры другого класса: *сознание, чувство, эмоция, воля, интеллект, мысль, представление, восприятие, идея, идеал, познание, вера* и т.д.

Настоящее исследование направлено на имя «сознание». В ходе рассмотрения его определения мы должны привести примеры проблем, связанных с ним наиболее тесно; это добавление обосновано тем фактом, что они станут яснее от его правильного определения. Эти проблемы отличаются от прочих, которые мы можем упоминать, но которые не могут быть решены теми же самыми средствами, какой бы тесной ни казалась их связь.

В ходе определения при помощи указанного выше инструментария мы можем обратиться в первую очередь к паре упомянутых последними терминов, а именно *сну* и *пробуждению*. В то время как *сон*, не сопровождаемый сновидениями, характеризуется отсутствием *сознания*, пробуждение от него является его появлением. Пробудившееся сознание может быть весьма разнообразным в его степени и содержании. Оно может быть настолько слабым, что не будет иметь никаких отличительных черт; оно может усилиться до любой стадии интенсивности; в то же время его свойства могут варьироваться настолько же, насколько различаются известные действия нашей психической жизни. Данный термин прекрасно применим ко всем колебаниям в определяемом. Он не указывает ни на какой особый тип психической активности; он обозначает только то, что ум ожил и работает по крайней мере в некоторых из его проявлений, а не приостановил свою деятельность или впал в сон. Необходимо указать на основания, по которым не стоит использовать это родовое имя для подразделения и классификации проявлений нашего ментального; для этой цели больше подходят другие слова.

В то время как первым и наиболее подходящим способом рассмотрения обычного понимания объема или значения сознания есть указание на сон и пробуждение, мы можем выявить дополнительные оттенки значения, указав на прочие случаи впадения в бессознательное состояние. Среди них сотрясение мозга, анестезия, упадок сил или истощение от перенапряжения, паралич мозга. Все они действуют так же, как и сон, указывая на привычное и понятное значение слова сознание негативным образом. Сейчас мы можем без опаски утверждать, что в психологическом словаре нет термина более несомненного, чем всеобъемлющий термин *сознание*.

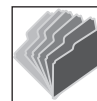


Легко привести множество прочих терминов, сопровождающих сознание, т.е. задающих условия, при которых ум предполагается бодрствующим, а не приостановившим свою деятельность. В силу своей понятности наиболее заметна из них пара *удовольствие и боль*. Естественно, существуют определенные теоретические трудности, до некоторой степени затуманивающие границы того, что ими обозначается; однако если не обращать на них внимания, обычно не возникает трудностей с их пониманием. Могут существовать подлинно психические процессы, сопровождаемые лишь малой толикой осознания, осуществляющие переход от одной мысли к другой на границе сознания, пусть даже не вполне бессознательные; однако удовольствие не есть удовольствие, если оно не осознается; мера удовольствия есть мера сознания, большее удовольствие или большая боль означают большее их сознание. В области *воли* эта пропорция неприменима без соответствующих поправок; она вполне подходит для сферы мотива или, другими словами, его *чувства*, но необязательна или не вполне годится для расхода энергии при выполнении решения. Один из наиболее известных и недвусмысленных процессов, называемый *привычкой*, является противоположным *сознанию* и в то же время приводит к психическим следствиям. Выводы в уме, будучи *сознательными*, задействуют определенную энергию или степень *чувства* или *воли*; они осознаются, если не вмешивается *привычка*, придавая им механический или автоматический характер.

Среди слов обыденной речи, не отягощенных двусмысленностью, находится термин «внимание». Эта счастливая свобода от сомнения несколько ограничивается употреблением термина для обозначения банальной интенсивности осознания, включая или исключая из нее добровольное побуждение. Тем не менее степень внимания есть мера степени сознания; полное невнимание будет означать полное отсутствие осознания какого-либо обстоятельства на протяжении некоторого времени.

Дальнейшие соображения относительно способа определения сознания будут сопровождаться рассмотрением центральных проблем, требующих решения.

**Объект сознания.** То, что наше распознавание так называемого внешнего, протяженного мира является модусом сознания, несомненно. Вопрос, приводящий к разногласию, касается в большей мере значения или сути того, что мы осознаем, но не определенности формы, посредством чего эта форма сознания помещается напротив различных форм, обозначаемых родовым именем *субъект*. Наша задача в настоящее время не включает в себя дополнительное рассмотрение общеизвестной противоположности объективных и субъективных качеств.



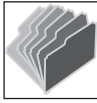
**Истины сознания.** Эта фраза имеет смысл только тогда, когда мы добавляем в обозначаемое *сознанием* нечто, не следующее из простого понятия бодрствования. То, что, бодрствуя, мы находимся в сознании и в полной мере, должно приниматься как несомненное. В этом мы не можем ошибаться. Даже широкое поле психических расстройств вряд ли включает в себя такой случай, что, будучи в состоянии какого-либо проявления сознания – удовольствия, боли, воления, мысли, кто-либо считает себя погруженным в глубокий сон. Предмет для работы появляется только тогда, когда возникают дополнительные вопросы, такие, как интуитивное знание об абсолюте за пределами наличного сознания. В тех случаях, когда подобные вопросы действительно появляются, их нужно рассматривать отдельно от термина *сознание*, под некоторым более специальным обозначением. Предполагаемая отчетливость сознания применима лишь к ограниченной области нашего строгого определения, за пределами которого стоит искать отчетливости другими способами.

**Сознание по сравнению с умом.** В то время как ум должно понимать объемлющим все разнообразие психических впечатлений, включая место, которое они занимают в умственной организации, будучи полностью в бездействии, или, иначе, когда они не выполняют никаких психических действий, термин *сознание* относится только к мгновениям психического бодрствования или ментальной действительности по отношению к наличным целям. Все постоянные товары, хранимые на психическом складе, нашли свой путь туда через период, когда они находились в *сознании*; они играют свою роль в экономике психического главным образом через возвращение к полному сознанию. Таким образом, сознание напоминает оформление сцены в театре в любой момент времени; это оформление представляется просто набором инвентаря из склада, хранящимся там со времени прошлых постановок или ожидающих будущих постановок.

Наш следующий заголовок также участвует в прояснении этого важного различия.

**Область сознания.** Это обозначение выражает черту сознания, не только жизненно важную саму по себе, но также разветвляющуюся на множество разнообразных вопросов. Взятое в любой отдельный момент, содержание сознания действительно ограничено довольно узкими рамками.

Как известно, область сознания ограничена единством исполнителя; это подмечено в поговорке о том, что каждому делу свой черед. Характеристики этой максимы имеют первостепенное значение в *психологии* и обычно раскрываются в связи с несколькими чувствами, в частности с чувствами протяженности – *зрением* и *тактильными* ощущениями. Упомянув это важное соображение, мы должны отметить, что каждое сознательное побуждение после того, как оно пре-



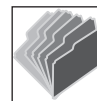
кратилось или угасла его сила, оставляет след. В этом следе или постоянном удержании коренится все разнообразие *памяти* или *ретенции*. Его действия имеют далеко идущие последствия; но более непосредственным образом пьесы сознания касается всегда готовое и непосредственное повторение того, что только что произошло в сознании по отношению к созидательной деятельности *ума* во времени. Можно сравнить этот процесс с тем, когда гончар отбрасывает куски глины, пока их не наберется достаточно для отдельного изделия. Наряду с сознанием любого отдельно взятого момента у нас есть некоторое количество недавних состояний, не попавших в сознание и постоянно возвращающихся хаотичным образом; эта хаотичность является лишь кажущейся, и условия, задающие это возвращение, вполне можно определить.

Узость области сознания является своеобразным ограничением человеческих возможностей по контрасту с нашим понятием *всезнания*. Строгость ограничения преодолевается определенной силой быстрого перехода, которая позволяет получить конструктивные результаты, включающие в себя несколько последовательных стадий сознательной репрезентации. Только благодаря одному этому обстоятельству мы затрудняемся сказать, сколько всего заключается в абсолютно мгновенном пучке *сознания*.

Практический вопрос большой важности, только упомянутый нами вскользь, заключается в том, чтобы описать наиболее подходящим для этой цели языком мотивы или побуждающие влияния сознания наряду с источниками предпочтительного внимания или концентрации среди соперничающих между собой элементов. Именно здесь мы должны принять решение относительно достоинств предложенной терминологии действий сознания, из которых следует дополнительный вопрос о взаимовлиянии трех великих частей, составляющих *ум субъекта*. Который из трех фундаментальных компонентов *ума* должен считаться основным в жизни нашей психики? Или каждый из них работает по очереди как первопричина?

**Сознание как неотъемлемая часть памяти.** Несомненно, что в общем случае для того, чтобы сделать постоянные приобретения или ассоциировать цепочки идей, такие последовательности должны начаться в сознании. Известным свойством удержания в памяти является то, что удерживаемое должно сначала полностью занять собой наше сознание, пускай оно будет большей или меньшей длительности, и чем более ярок огонь сознания, тем быстрее и более связно эти моменты нарастают. В отношении сказанного как общего принципа не может быть никаких сомнений – это основа и главное условие любого успешного обучения. На него также влияет физическое состояние нервной системы, которая может быть предрасположена к возбуждению, а не постепенному наращиванию ассоциативных связей. Од-



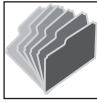


нако сказанное не влияет на главный тезис. Кажущиеся исключения относятся к совершенно другому виду. Несомненно, существуют явления, стимулирующие связный рост и за пределами области сознания. Обратившись к процессу взросления ребенка, когда рефлекторная и спонтанная активности закрепляются их повторением, мы указываем на внесознательную область нашего обучения. Без сомнения, в данной области последовательные действия уже существуют и в дальнейшем просто закрепляются. Однако могут ли два различных типа движения выводиться из сознания – совершенно иной вопрос. Не существует ничего, подталкивающего нас к предположению, что это каким-либо образом осуществимо. Когда мы имеем дело с впечатлениями, доставляемыми нам различными органами чувств, их объединением в группы и последовательности, мы должны без малейших колебаний заявить, что такие объединения начинаются в сознании, а их скорость определяется интенсивностью осознания.

Таким образом, перед нами встает проблема, следующая из имени *сознание* и не подпадающая ни под какое иное имя: существуют ли значимые психические изменения, возникающие в пробелах нашего сознания, таких, как сон или временное отвлечение от предмета рассмотрения?

**Непосредственные физические условия для сознания.** Обширная проблема связи ума и тела, глубина и различные ответвления которой намного превосходят наши самые оптимистические надежды на будущие исследования, приобретает более обозримый вид, если мы ограничим рассмотрение собственно сознанием, как мы его определили. Несмотря на то что переходы от сна к бодрствованию, от слабого к интенсивному сознанию еще не поняты во всей мере, они все же связаны с ощутимыми и вполне объяснимыми физическими изменениями, которые поддаются объяснению в явном виде и обладают важнейшим практическим значением. Из этих связей между психическим и физическим мы выводим значимые следствия относительно великой проблемы соединения ума и тела во всем ее охвате. Известно, что сопровождение телесных *выражений* состояниями *чувствования* поддается измерению и является основой для срабатывания нервных связей, которые, по-видимому, требуются для сознания.

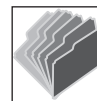
**Рефлекторные действия и сознание; животный автоматизм.** Как правило, классификацию рефлекторных действий мы начинаем с тех, в которых сознание не задействовано, например дыхание, и заканчиваем теми, где оно участвует и является до определенной степени управляющим, но необходимым только отчасти, например при отдергивании руки от раскаленной поверхности. Термины «бессознательный» и «сознательный» здесь применимы в самом строгом и уместном смысле, а отсылка к ним позволяет закрепить типичные значения слов. Они дополнительно показывают связь сознания с нашей



подлинно произвольной деятельностью во всем их разнообразии. Действия теряют свои подлинно произвольные черты в двух крайних случаях или переходных ступенях: с одной стороны, с их затенением в *рефлекс*, а с другой – в их переходе в *привычность*. В обоих случаях они до определенной степени лишаются сознательных свойств, как это видно из того, что они уступают место другим членам области сознания.

Новая черта проблемы *сознания* проявляется в том случае, когда мы задаемся вопросом: являются ли животные автоматами? Вполне можно предположить, что нервная система во всей ее сложности и согласованности могла бы производить все действия, к которым способны животные, без участия сознания, равно как и при его наличии. Очевидная трудность заключается в том, что в нашем собственном опыте мы имеем два типа психической активности, один из которых сопровождается сознанием, а другой – нет, и что животные могут быть как наивысшего, так и низшего типа. В нашем случае сознание есть условие для приобретения способностей, благодаря ему мы учимся на опыте, а не являемся просто машинами, выполняющими врожденные и повторяющиеся действия. Низшие животные таким же образом учатся на опыте, и было бы несправедливо, если бы мы предположили, что их приобретенные способности не связаны с сознанием. В нашем случае усиление сознания ускоряет постоянные впечатления и позволяет обучаться. То же самое предполагается и применяется при нашей дрессировке животных. Таким образом, мы, кажется, приписываем им такое же сознание, каким обладаем сами, в различных вариациях, частично улавливаемых в отсылке к вариациям в нашем собственном сознательном опыте. Мы видим в собаке те же порывистые смены внимания, те же скачки в осознании цели с той же способностью к восстановлению, какие обнаруживаем в себе в схожих условиях. Если мы колеблемся в применении к животным предложенного различия, то нам нужно принять иной язык для описания происходящего в их психике.

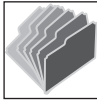
Доказательства в пользу сознания у животных могут быть объединены под следующими заголовками: (1) Строение мозга, чрезвычайно сильно напоминающее наше, особенно у высших животных; данное сходство не прерывается вплоть до беспозвоночных, у которых строение мозга радикально иное. (2) Явное проявление тех же действий, которые типичны для людей в тех же обстоятельствах, когда мы испытываем удовольствие или боль. (3) Действие тех же обстоятельств на преследующее и избегающее поведение, т.е. той же произвольной деятельности, к которой они приводят в человеческих существах. Совокупная сила этих доказательств всегда считалась доводом в пользу сознания у животных по противоположности с меха-



низмом, воплощенным в рефлекторной деятельности, несмотря на любую предполагаемую степень возникающих здесь сложностей.

В то же время, как мне кажется, даже более сильным, чем любое из этих доказательств, будет соображение, от которого мы выше отвлеклись, касающееся абсолютной необходимости сознания для овладения чем-либо. Ни один факт нашего строения не является более несомненным; отказ от применения этого соображения к прочим существам, способным к обучению, является необоснованным и неоправданным. Вместо того чтобы облегчить наши трудности в области *психологии животных*, он усиливает их до самой крайней степени. В качестве доказательства упомянутый факт венчает три предшествующих доказательства по аналогии и превосходит их по убедительности.

Природа и степень сознания любого отдельно взятого представителя племени животных часто является вопросом умозрительного любопытства. Даже человеческие существа, пытаясь проникнуть в сознание друг друга, часто ошибаются и никогда не уверены в полной мере, стоит ли за теми же признаками то же самое – то же состояние сознания. Именно таково наше фундаментальное затруднение в случае с животными. Применяя все четыре типа указанных выше признаков, мы должны сделать вывод о состоянии и степени сознания в каждом конкретном случае. Возможно, наиболее сильным доводом в пользу участия сознания является то, на чем мы главным образом и сосредоточились, а именно: обучаемость. Страстность выражения и произвольное преследование цели или избегание опасности проявляются как в низших, так и в высших порядках – в насекомом или рыбе так же, как и в млекопитающем. Значение этих признаков, взятое по отдельности, крайне неясно и запутанно. Они сопровождают низшие виды мозга настолько же, насколько и высшие. Однако вполне ощутимое различие предоставляется нам в том, способен ли и в какой степени индивид к великим основаниям разумности – *различению и обучению*. Тип, если не степень сознания, должен проявляться все ближе и ближе к человеческому виду, согласно тому как нарастают эти способности. Несмотря на весь наш разнообразный опыт насыщенности сознания или его бодрствования, мы можем быть неспособны к точному представлению природы и степени низших беспозвоночных, обладающих чувствами и реагирующих как по выражению, так и при помощи движения на внешние раздражения. Следствием из сказанного будет невозможность указать на ту точку в развитии животных, в которой впервые проявляется сознание. Мы можем вполне справедливо считать его присутствующим тогда, когда выразительные реакции и произвольное преследование цели сопровождаются самой малой долей обучаемости. Будучи вопросом эволюции или развития, генезис сознания, по всей видимости, неразрешим нашими

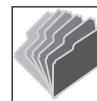


средствами. Он сопоставим с вопросом о том, чему отдать преимущество – *выражению* или *волеию*, что само по себе находится недалеко от другого неразрешимого вопроса об источнике или исходной точке нашего *рефлекторного приспособления*: является ли оно следствием пережитых на опыте или приобретенных действий; в этом случае оно уже будет предполагать некоторую стадию сознания, а не находиться перед ним.

**Сознание и самосознание.** Термин «самосознание» открывает обширное поле для дискуссий, равно как и включается во многие из наших глубочайших разногласий. Имя *сознание* само по себе, как ясно из предыдущего, является вполне понятным и свободным от двусмысленности. Добавление к нему приставки «само» полностью меняет положение дел. Самость, взятая по отдельности, обладает различными значениями; то же разнообразие вторгается в ту область, в которую она входит как составная часть.

Когда к *сознанию* присоединяется характеристика, это делается для того, чтобы ограничить его общность или объем до какого-либо предела: во время бодрствования мы обычно обладаем набором вещей, представленных перед нами или которые доступны нашему взгляду, в то время как только одна или очень небольшое их число могут находиться в сознании в один и тот же момент времени. Вполне возможно определить, на который из разнообразных единомоментных стимулов обращено внимание, а какие из них игнорируются; о нем мы скажем, что он находится в сознании, а об остальных – что они бессознательны. Покуда эти противоположности имеют простую, недвусмысленную форму, их соединение со словом *сознание* не убавляет понятность языка. В моменты рассеянности мы не осознаем то, что находится перед нами, или то, что мы слышим. Мы даже можем не осознавать физическую боль. Но все же можно представить себе значительные трудности, связанные с таким ясно определенным пониманием имен «сознание» и «бессознательное». Другим примером допустимой характеристики сознания при помощи обращения к узкой теме является риторическая фраза «*mens conscia recti*»<sup>3</sup>. В ее использовании нет ничего запутывающего, хотя было бы предпочтительней использовать более подходящие слова. Правота чьих-либо намерений и поведения не познается во всей своей полноте за одно мгновение сознания; она требует сравнения с множеством подобных случаев, следовательно, предпочтительней было бы использовать слово «знание». Схожие возражения можно выдвинуть при описании сознания как проверки или свидетельства наших интуиций. У нас могут быть интуиции, и в них может быть какая угодно степень правоты, но в ней нельзя удостовериться при помощи одного единственного

<sup>3</sup> В своем праве / в сознании своего права (Овидий). – Прим. перев.



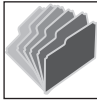
момента осознания, а сознание не может нести с собой память без того, чтобы выйти за свои ограниченные пределы.

Когда же в качестве содержания сознания мы вводим термин «самость», трудность становится поистине непреодолимой. Для прояснения мы должны немедленно указать, какие из общепринятых значений этого термина мы имеем в виду. В обычной речи под выражением «самосознательный» принято понимать то, что кто-либо примеряет к себе свою собственную значимость, признание или заслуги в глазах других. Тщеславный человек, когда он этому чувству потворствует и особенно когда привлекает внимание других, зовется самосознательным<sup>4</sup>. В противном случае самооценка может принять форму нездорового самоуничижения, например когда вследствие некоего действия или обстоятельства, производящего плохое впечатление на окружающих, индивид негативно оценивает сам себя. Эти два крайних случая представляют ту форму *самости*, которая в современной речи обычно присоединяется к имени *сознание*. Более ясное значение самости как включающее в себя все наши жизненные интересы или коллективные ценности лучше передается связкой «своекорыстный».

Общепризнанно, что слово *сознание* объемлет мир *объектов* точно так же, как и *субъекта*. Противоположность этих двух модусов имеет такой характер, что необходимо пояснять, если упоминается именно один, а не другой. Таким образом, когда мы намеренно опускаем отсылку к объекту, мы можем обозначить остаток как *самосознание*. (Мне кажется, здесь не нужно указывать на старый способ использования *сознания* как *интроспекции* или источника нашего знания об *уме*, соответствующего *наблюдению* для мира *объектов*.) Однако, имея перед собой такое количество общепринятых значений слова *самость*, подобное использование данного термина неуместно и неудовлетворительно, хотя и не лишено отдельных достоинств. Иногда говорят, что все наши чувства, познавательные способности и воления отсылают к самосознанию, что является просто иным способом сказать об их принадлежности уму субъекта. Использовать имя «сознание» подобным образом означает перешагнуть за его пределы, так как оно является просто выражением преходящих стадий нашего психического бытия, и смешать его с всеобщностью *ума*, которая больше, чем любая из подобных стадий, в сто тысяч раз.

Мне неизвестны другие настолько же общепринятые способы использования рассматриваемого соединения, помимо двух рассмотренных. До этих пределов оно лишено двусмысленности настолько, что его можно использовать и в повседневной жизни, и в психологической теории.

<sup>4</sup> В русском языке этому соответствует слово «самодовольный». – Прим. перев.



Когда же «самосознание» подается как высший факт из всех, что нам известны, или как наш «ключ к последней природе всего существующего в его целостности», делается новый шаг в сторону расширения его значения. Ни одна из двух составных частей связки не кажется способной содержать этот исключительно важный вопрос. Конечный предел или граница самости как тотальность нашего бытия – ума и тела; касательно этой самости мы можем сказать, что она сознает в смысле знания, что будет гораздо более уместным термином для такой всеохватной агрегации. Предположим, что мы используем «самознание» для целей нахождения последней природы существования. К чему можно приравнять это употребление? Только к тому, что человечество, взятое как целое – ум и тело, настолько мощная созидательная и всевластная сила во вселенной, что делает недостаточным и неудовлетворительным объяснение происхождения мира, какой он есть. Или, короче, это будет просто повторением антропоморфного объяснения природы, господствовавшего такое долгое время. Предпочтение пары само–сознание будет означать уменьшение силы такого утверждения, так как *сознание*, будучи ограничено наличным и ускользающим, является более узким термином, нежели *знание*, которое охватывает весь постоянный массив или собрание всего, что когда-либо присутствовало в сознании.

Быть может, «самость» в рассматриваемой связке должна относиться только к *уму*, т.е. уму как чистому и абстрактному опыту, отличному от тела, пусть и неотделимому от него. Тело при таком подходе представляет собой лишь помеху, не обладая ничем, что бы соответствовало предполагаемой творческой активности мира: действующая продуктивность ума постулируется без всякой связи с физическим строением, в отсутствие которого мы на самом деле не могли бы переживать *ум* как отдельную сущность.

В термине, рассмотрением которого мы занимались, я не вижу никакой пользы в великой борьбе за выяснение источников нашей веры в Реальность как противоположность Видимости. И действительно, если я правильно понимаю современное состояние этой борьбы, его нечасто делают ведущим. Если же его сделают таковым, то между противоборствующими сторонами должно существовать четкое согласие относительно его точного определения; я бы даже сказал, что его стоит заменить любым другим фразеологизмом, менее укорененным в двусмысленности.

В заключение нашего рассмотрения основного тезиса, определения и проблем сознания достаточно привести обоснованное соображение: критическое исследование целостного образования «самосознание» уже намного выходит за границы области психологии.

*Перевод с английского А.А. Веретенникова*



## СОВРЕМЕННЫЙ ДУАЛИЗМ: ПОПЫТКА ЗАЩИТЫ

## CONTEMPORARY DUALISM. AN ATTEMPT AT DEFENSE

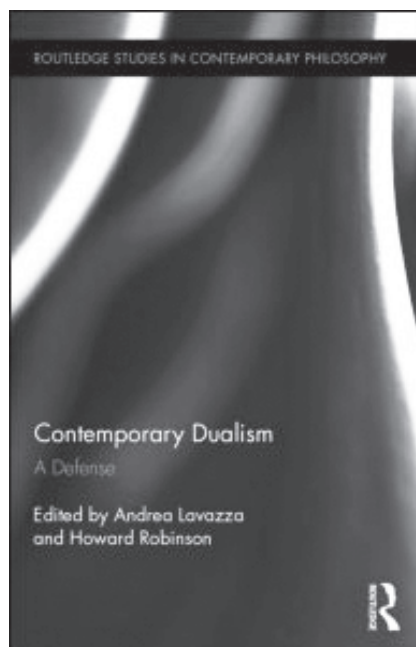
**Игорь Гарибович Гаспаров** – кандидат философских наук, старший преподаватель Воронежской государственной медицинской академии им. Н.Н. Бурденко. E-mail: gasparov@mail.ru

**Igor Gasparov** – Voronezh Medical State Academy.

(Рецензия на книгу: Contemporary Dualism: A Defense; A. Lavazza and H. Robinson (eds.). Routledge, 2014. 292 p. ISBN 9780415818827.)

За последние пять лет наряду с рецензируемым ниже изданием вышло в свет несколько сборников, посвященных защите современного дуализма [Koons and Bealer, 2010; Baker and Goetz, 2011]. Несмотря на то что физикализм по-прежнему занимает доминирующую позицию среди современных работ по проблемам сознания, начиная с 1990-х гг. можно говорить если не о ренессансе дуализма, то по крайней мере о его некотором оживлении. Другими словами, дуализм вернулся в философскую повестку дня и обсуждается как серьезная альтернатива материализму. Основной причиной этого стала, по мнению сторонников дуализма, принципиальная неспособность материализма дать адекватное объяснение таким характеристикам психических явлений, как интенциональность и субъективность. При этом сам дуализм также претерпел значительные изменения и обновил свой облик. Во-первых, среди философов возросла неудовлетворенность физикалистскими объяснениями сознания. Это подтверждают последняя книга Томаса Нагеля [Nagel, 2012] и уже давно известная критика современной научной картины мира Питером Ангером [Unger, 2005]. Во-вторых, на рынке философских идей появились новые разновидности этой позиции – эмерджентный дуализм и другие вариан-

ты некартезианского дуализма, которые в целом можно охарактеризовать как натурализацию дуализма. В-третьих, в своей аргументации сторонники современного дуализма опираются не только на традиционные априорные аргументы против материализма, известные со времен Р. Декарта, Дж. Беркли и Г. Лейбница, но и на данные современных эмпирических исследований. В-четвертых, сами





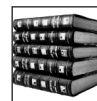
эмпирические исследователи, – правда, скорее косвенно, нежели напрямую, – стали более лояльно относиться к дуализму. Наиболее яркими примерами могут служить недавние заявления Кристофа Коха [Koch, 2012] и Джулио Тонони [Tononi, 2012]. Рецензируемый сборник, отражающий перечисленные выше тенденции, включает 13 самостоятельных статей и введение, написанное издателями Андреа Лавацца и Ховардом Робинсоном. Статьи распределены по четырем разделам: «Границы материализма», «Дуализм и эмпирические исследования», «Картезианский дуализм субстанций» и «Некартезианский дуализм».

В первом разделе читатель может познакомиться с работами, посвященными критике материализма в философии сознания, и попытками показать, что дуализм предлагает более удачное объяснение сознания. Раздел представлен статьями Дэвида Лунда, Уве Майкснера и Андреа Лавацца. Наиболее оригинальной является работа последнего “Problems of Physicalism Regarding the Mind”, в которой автор приводит аргументы в пользу того, что физикализм в отношении сознания в союзе с эволюционистским пониманием происхождения мозга имплицитно контингентный характер законов логики и допускают существование несоизмеримых между собой концептуальных схем, что в свою очередь ставит под вопрос универсальный характер не только логики, но и морального закона.

Второй раздел посвящен соотношению дуализма и эмпирических исследований. Здесь значительный интерес представляет работа Риккардо Манзотти и Паоло Модерато “Neuroscience: Dualism in Disguise”. В ней авторы представляют аргументы в пользу того, что современная нейронаука, несмотря на весомый

прогресс в области исследований механизмов работы клеток головного мозга, мало продвинулась в понимании природы самого сознания по той причине, что руководствовалась в этих исследованиях противоречивыми посылами, по существу исключая возможность решения проблемы сознания. С одной стороны, нейронаука предполагает, что физическая реальность по определению не может включать в себя какие-либо ментальные свойства (интенциональность или субъективность), а с другой стороны, утверждает, что сознание, обладающее всеми этими свойствами, каким-то образом способно находиться внутри одного полностью материального объекта – центральной нервной системы. По мнению авторов, именно одновременное принятие этих двух идей заставляет современных нейрочеловеков вводить в научный оборот различные сущности с неопределенным онтологическим статусом, как, например, «информация» или «ментальная репрезентация», что является замаскированным признанием дуализма. Авторы объясняют такую стратегию реакцией современной нейронауки на неспособность примирить собственные постулаты, приведенные выше. В целом, я думаю, что итальянские авторы правы в своих замечаниях о состоянии современной науки. Однако среди нейрочеловеков есть и такие, кто эксплицитно признает реальный онтологический статус за психическими сущностями, открыто принимая дуализм. Хорошим примером такой позиции является «материальный дуализм» Джона Смитиса [Smythies, 1994]. Согласно этой позиции, психические репрезентации вопреки мнению Декарта не лишены протяженности, а образуют собственное трехмерное пространство, дополняющее наше привычное физиче-





ское пространство, также имеющее три измерения.

В другой статье известный специалист в области квантовой механики Генри Пирс Стэпп, развивая «квантовую теорию сознания», предлагает оригинальный ответ на вопросы, поднятые экспериментами Бенджамина Либета, относящимися к возможности свободного выбора субъектом своих действий. Стандартная интерпретация этих экспериментов исходит из следующего. Индивид не может свободно инициировать свое действие, поскольку если верно, что причина всегда предшествует следствию, то свободное решение должно быть осознано индивидом раньше, чем мозг начнет посылать импульсы, задействующие члены, приводящие это решение в исполнение, тогда как данные, полученные Либетом в ходе экспериментов, кажутся, свидетельствуют об обратном. Согласно этим данным, сначала в мозге возникает потенциал готовности и лишь затем индивид фиксирует свое сознательное решение совершить действие. Стэпп полагает, что фиксация в мозге потенциала готовности должна интерпретироваться как возникновение в сознании репрезентаций возможных последствий совершаемого действия, которые всегда предшествуют принятию решения о его совершении или несовершении.

Наконец, в последней статье данного раздела австрийский философ Мартина Фюрст из Грацкого университета предлагает оригинальную трактовку «стратегии феноменальных концептов», которая является одним из наиболее популярных ответов на антифизикалистские аргументы в стиле Фрэнка Джексона. Фюрст приводит детальный анализ литературы, посвященной «стратегии феноменальных концептов», и приходит к выводу, что сторонникам

физикализма не удастся удовлетворительно объяснить, почему идеальный физик Мэри узнает некие новые факты о мире после того, как покинет черно-белую комнату, несмотря на то что она и до этого обладала знанием *всех* физических фактов о нашем мире. Критика Мартины Фюрст направлена прежде всего на концепцию Дэвида Папино. Она полагает, что ему не удастся дать адекватное объяснение того, как работают феноменальные концепты, и приводит аргументы далее в пользу того, что сами феноменальные концепты предполагают существование нефизических референтов.

Третий раздел посвящен картезианскому дуализму. Ричард Суинберн в своей работе “What Makes Me Me? A Defense of Substance Dualism” приводит аргументы в поддержку отождествления личности с простой нематериальной субстанцией, которая лишь контингентным образом соединена с тем или иным физическим телом. В статье под заголовком “Naturalism and the Unavoidability of the Cartesian Perspective” Ховард Робинсон отстаивает тезис, что если верна идея о каузальной замкнутости физического, то такие науки, как химия, биология, психология, не могут иметь самостоятельной функции. Робинсон выступает в пользу тезиса, что перспектива первого лица не может быть полностью устранена из науки и что наука зависит от нее. Ральф Уолкер выдвигает тезис, что ментальное является нередуцируемой сферой реальности, поскольку она в отличие от физических объектов чувствительна к объективным требованиям разума. Этот способ воздействия принципиально отличен от каузальных отношений, присутствующих в материальном мире. Интересно, что подобное традиционно кантианское различие между причинами и доводами Уолкер использует



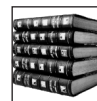
для обоснования тезиса о картезианском дуализме субстанций, который он защищает во второй части своей статьи.

Последний, четвертый раздел книги посвящен некартезианским формам дуализма. Здесь представлены четыре автора, которых объединяет прежде всего то, что они не разделяют ни тезис материализма, ни тезис картезианского дуализма. На мой взгляд, наиболее интересной работой этого раздела является статья ныне уже покойного профессора Даремского университета Эдварда Джонатана Лоу “Why My Body is not Me: The Unity Argument for Emergentist Self-Body Dualism”. Он приводит оригинальный аргумент в пользу того, что личность должна быть признана онтологически базовой категорией. Суть аргумента состоит в том, что если мысль не может существовать без субъекта, то *моя* мысль не может существовать без *меня*. Лоу пытается показать, что ни один материальный объект – будь то целый живой организм, или его часть – не является необходимым для существования всех моих мыслей. Следовательно, делает он вывод, я не могу быть тождествен никакому материальному объекту. И поскольку антиреализм относительно моего существования кажется сомнительным, личность, или субъект мысли, должны быть признаны в качестве фундаментального и ни к чему несводимого элемента реальности. При этом Лоу отказывается отождествлять личность с картезианской душой, аргументируя тем, что ее природа находится за пределами его понимания. Позиция Лоу сродни позиции Линн Раддер Бейкер [Baker, 2000] с тем различием, что Бейкер позиционирует себя как сторонница ненатуралистического материализма, в то время как Лоу говорит об эмерджентном Я (Self) и отказывается отождествлять

свою позицию с материализмом. На мой взгляд, концепции и Бейкер, и Лоу страдают одним и тем же недостатком: трудно понять, как личности, если они представляют собой онтологические базовые сущности, соотносятся с другими привычными нам «естественными видами», такими, как физические частицы или живые организмы.

Оставшиеся работы наглядно демонстрируют разнообразие позиций и тем, объединенных лейблом некартезианского дуализма. Чарльз Талиаферро в статье “Promise and Sensibility of Integrative Dualism” аргументирует в пользу трех связанных между собой тезисов. Во-первых, перспектива первого лица концептуально предшествует перспективе третьего лица. Во-вторых, естественной установкой для дуалиста является неразрывная связь между душой и телом, а не отделение одного от другого. В-третьих, вопреки общепринятому мнению для дуалиста взаимодействие между душой и телом не является такой загадкой, как это представляется физикалистам.

Уильям Хаскер в статье “The Dialectic of Soul and Body” обращается к проблемам томистского дуализма. Он утверждает, что хотя восходящая к св. Фоме Аквинскому концепция человека как единства души и тела обладает важными преимуществами по сравнению с картезианским дуализмом, ни она, ни ее современные модификации (в качестве примера Хаскер рассматривает концепцию Джеймса Портера Морлэнда) неспособны ответить на вызовы современной науки в лице дарвиновской теории эволюции. В связи с этим Хаскер выступает в пользу эмерджентного дуализма, который, по его мнению, сохраняет достоинства томистской концепции личности, но лишен ее недостатков, в частности он хорошо сочетается с современными пред-



ставлениями об эволюции. Статья Дэвида Скрбины посвящена двухаспектной теории сознания. По мнению автора, эта позиция удачно объединяет преимущества традиционно конкурирующих между собой теорий – дуализма и материалистического монизма.

Если оценивать сборник в целом, то следует сказать, что благодаря разнообразию представленных

в нем авторов и тем читатель получит разностороннее представление о современном дуализме как живом и интересном явлении, знакомство с которым может быть полезным не только для тех, кто так или иначе отождествляет себя с этой позицией, но и для тех, кто хочет взглянуть на проблематику современной философии сознания с точки зрения, отличающейся от ортодоксальной.

### Библиографический список

Baker L.R., 2000 – *Persons and Bodies: A Constitution View*. Cambridge, UK : Cambridge University Press.

Baker M. and Goetz S., 2011 – *The Soul Hypothesis*. L. ; N.Y. : Continuum Press.

Koch C., 2012 – *Consciousness: Confessions of a Romantical Reductionist*. Cambridge, MA : The MIT Press.

Koons R. and Bealer G. (eds.), 2010 – *The Waning of Materialism*. N.Y. : Oxford University Press.

Nagel T., 2012 – *Mind and Cosmos: Why the Materialist Neo-Darwi-*

*nian Conception of Nature is Almost Certainly False*. N.Y. : Oxford University Press.

Smythies J., 1994 – *The Walls of Plato's Cave: Science and Philosophy of Consciousness, Brain and Perception*, Avebury.

Tononi G., 2012 – *Phi: A Voyage from the Brain to the Soul*. Pantheon Books.

Unger P., 2005 – *All the Power in the World*. Oxford : Oxford University Press.

# ЮБИЛЕЙ

## АЛЕКСАНДРА ЛЕОНИДОВИЧА НИКИФОРОВА

В конце апреля этого года член редколлегии нашего журнала, доктор философских наук, главный научный сотрудник сектора социальной эпистемологии Института философии РАН Александр Леонидович Никифоров отметил свой 75-летний юбилей.

Александр Леонидович был и остается не только квалифицированным специалистом в области логики, методологии и философии науки, но оригинальным и глубоким мыслителем, сфера исследовательских интересов которого простирается далеко за пределы его официальной квалификации. За последние годы Александр Леонидович провел ряд интереснейших исследований в области философии истории, показав, например, как именно в работе историка в отличие от работы естествоиспытателя описание оказывается неразрывным образом связано с объяснением. Его переоценка целей и задач современных науки и техники, согласно которой они не были направлены на поиск истины, а ориентировались преимущественно на то, чтобы сделать существование человека наиболее комфортным, стала неожиданной для целого ряда известных специалистов в этой области. В последнее время его критика и переосмысление философии либерализма вызывает острые дискуссии и живой интерес как у сторонников, так и у противников этого мировоззрения, затрагивающего практически все сферы жизни общества.

Александр Леонидович – замечательный полемист, на его докладах всегда аншлаг, а дискуссии по поставленным им вопросам зачастую продолжаются гораздо дольше отведенного регламентом времени. И несмотря на ту едкую критику, которой он порой подвергает работы своих коллег, мало к кому в Институте философии относятся с таким уважением и теплом, как к нему.

Александр Леонидович был и остается прекрасным лектором и популяризатором научного знания. Его «Популярные лекции по логике и теории аргументации», выложенные на сайте Института философии РАН, открытые выступления и участие в интернет-проектах расширили и продолжают расширять его популярность среди многочисленной интернет-аудитории, интересующейся философией.

Редколлегия журнала от всей души поздравляет Александра Леонидовича с юбилеем и желает ему здоровья, дальнейшей плодотворной работы и как можно больше тех положительных эмоций, которые он всегда дарит своим студентам, коллегам и друзьям.

---

# Памятка для авторов

## 1. ОБЩИЕ РЕКОМЕНДАЦИИ

При написании статей рекомендуется учитывать профиль издания и строить содержание и форму статьи применительно к одной из рубрик журнала. Предлагаемые материалы должны являться не опубликованными ранее научно-философскими текстами, обладающими актуальностью и новизной. Объем любого материала – до 1 а.л.

## 2. ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ:

- материалы принимаются по электронной почте в формате doc (шрифт – Times New Roman, размер – 12, междустрочный интервал – одинарный);
- на первой странице статьи должно быть: русское и английское названия текста, русскоязычные и англоязычные данные об авторе (ФИО, ученая степень, должность и место работы, e-mail), русскоязычная и англоязычная аннотации с ключевыми словами (англоязычная аннотация должна быть расширенной – около 1,5 тыс. знаков с учетом пробелов);
- сноски размещаются в низу страницы, сквозная нумерация;
- ссылки на литературу даются в тексте статьи в квадратных скобках – фамилия автора и год (если надо, номер страницы): [Сидоров, 1994: 25]. После текста на последней странице прилагается библиографический список в алфавитном порядке, где для каждой ссылки сначала приводится ее сокращенное обозначение (которое в тексте давалось в скобках, но уже без указания статьи) и рядом через тире полные выходные данные: Сидоров, 1994 – *Сидоров И.И.* Название книги. Город, год;
- в конце статьи также следует предоставлять библиографический список на латинице, в котором выходные данные русскоязычных источников будут транслитерированы по правилам научной транслитерации русского языка: [http://en.wikipedia.org/wiki/Scientific\\_transliteration\\_of\\_Cyrillic](http://en.wikipedia.org/wiki/Scientific_transliteration_of_Cyrillic);
- к тексту статьи следует прилагать фотографию автора.

### В ССЫЛКАХ ОСТАВЛЯТЬ ТОЛЬКО СЛЕДУЮЩИЕ СОКРАЩЕНИЯ:

- нем., англ., амер., греч., лат. – и др. языки;
- пер. – перевод;
- соч. – сочинение, сочинения;
- кн. – книга;
- Т. – том;
- Ч. – часть.

### СОКРАЩАЮТСЯ НАЗВАНИЯ ГОРОДОВ (В ССЫЛКАХ):

М., Л., СПб. – Москва, Ленинград, Санкт-Петербург.  
L., P., N.Y., F.a.M. – Лондон, Париж, Нью-Йорк, Франкфурт-на-Майне.  
Сначала идут русские названия (если есть), затем – названия на иностранном языке. Автор, название, место и год издания – L., 1965; M., 1995. Работы отделяются друг от друга точкой с запятой (;). Если в библиографию включается статья, то книга или журнал, в которых она напечатана, приводится через знак //. Названия журналов – без кавычек, без курсива и без сокращений.  
*Иванов В.С.* Либерализм Ф. Хайека. М., 1997; *Popper K.* Open Society. V. 1. Oxford, 1956.

## 3. ПОРЯДОК ПРИНЯТИЯ СТАТЬИ

Материалы рассматриваются в течение трех месяцев двумя независимыми рецензентами и далее редколлегией, которая принимает окончательное решение о публикации.

## 4. МАТЕРИАЛЫ ПРИНИМАЮТСЯ ПО АДРЕСУ:

journal@iph.ras.ru

5. По желанию автора ему может быть представлен мотивированный отзыв в случае отказа редакции журнала от публикации его статьи.

6. С автором текста, одобренного редколлекцией, заключается договор о передаче ООО «Издательство «КноРус» исключительных прав на его публикацию сроком на 1 год.

**За публикацию материалов плата не взимается и гонорар не выплачивается.**

### Information for Contributors

All manuscripts are submitted by e-mail and must be sent to: [journal@iph.ras.ru](mailto:journal@iph.ras.ru).

#### Requirements for articles and book reviews:

Please, use DOC file type. Page size: A4. Font: Times New Roman, size 12. Do not double-space. Author information, abstract and key words must be sent in a separate file while another separate file containing the text must be devoid of personal data and prepared for the blind peer review. Please, use notes on the page they appear in the text. The list of references must follow the manuscript. In the text we prefer the references to be of the following style: author's last name (date), section or page(s).

The article's recommended size is 3000–6000 words.

### Review and Publication Time

Evaluation time for manuscripts of articles by blind peer reviewers is up to 3 months. All evaluated materials can be revised by the editorial board within 3 months after evaluation. Publication time for approved materials is within 3 months. Total publication time is up to 9 months.

Unsolicited book reviews are invited. The standard size of a review is 1 thousand words.

---

## Подписка

Уважаемые коллеги. Наш журнал распространяется как в розницу, так и по подписке. Журнал выходит ежеквартально. Годовая подписка состоит из 4 номеров.

Кроме того, в настоящее время альтернативную подписку журнала осуществляют: «Интерпочта» (Москва), «Информнаука» (Москва), «Красносельское агентство «Союзпечать»» (Москва), «Пресс Инфо» (Казань).

Читатели могут также получить любое количество номеров журнала (от 1 до 4 в год), лично обратившись в редакцию.

Индекс в каталоге Роспечати: **46318**

### Адрес редакции:

119991, Москва, Волхонка, 14/1, стр. 5  
Институт философии РАН  
Телефон: (495) 697-9576  
Факс: (495) 697-9576  
Электронная почта:  
journal@iph.ras.ru

### Адрес издательства:

117218, Москва, ул. Кедрова, д. 14,  
корп. 2  
ООО «Издательство «КноРус»  
Тел./факс: +7 (495) 741-4628  
Электронная почта: welcome@knoirus.ru

Более подробную информацию см. на сайте журнала <http://iph.ras.ru/journal.htm>

## Subscription Information

All potential subscribers from outside the Russian Federation or CIS countries must contact the editor: [journal@iph.ras.ru](mailto:journal@iph.ras.ru).

Current rates for institutional subscribers: 270 USD per year, 80 USD per issue; for individual subscribers: 220 USD per year, 60 USD per issue.

For more information please see the journal's web page: [eng.iph.ras.ru/journal.htm](http://eng.iph.ras.ru/journal.htm).

---

### Вниманию подписчиков

Журнал «Эпистемология и философия науки» прошел перерегистрацию в Агентстве «Роспечать» и с 1 января 2015 г. выходит под названием «**Epistemology & Philosophy of Science / Эпистемология и философия науки**». Все обязательства по подписке сохраняют свою силу, подписной индекс не меняется. С июня 2014 г. журнал входит в международную базу данных «Philosophy Documentation Center», которая будет обеспечивать open access журнала.

К публикации принимаются статьи на русском и английском языках.

**Epistemology & Philosophy of Science / Эпистемология  
и философия науки. 2015. Т. XLIV. № 2**

Главный редактор чл.-корр. РАН *И.Т. Касавин*  
Заместитель главного редактора д-р филос. наук *И.А. Герасимова*  
Ответственный секретарь канд. филос. наук *П.С. Куслий*

Подписано в печать 27.05.2015  
Формат 60 × 100 <sup>1</sup>/<sub>16</sub>. Печать офсетная. Бумага офсетная.  
Печ. л. 16,0. Тираж 800 экз. Заказ №

ООО «Издательство «КноРус»  
*Адрес:* 117218, Москва, ул. Кедрова, д. 14, корп. 2  
*Тел./факс:* +7 (495) 741-4628  
*E-mail:* welcome@knorus.ru

*Адрес редакции:* 119991, Москва, Волхонка, 14/1, стр. 5  
Институт философии РАН. *Тел.:* (495) 697-9576  
*Факс:* (495) 697-9576. *E-mail:* journal@iph.ras.ru