

К 5395571

ХИЛАРИ ПАТНЭМ

ФИЛОСОФИЯ

СОЗНАНИЯ

5395571



ХИЛАРИ ПАТНЭМ

(Род. в 1926 г.)

“На деле нас интересует, как говорил Аристотель, не материя, а форма. Вопрос не в том, что есть материя, а в том, что есть наша интеллектуальная форма? И чем бы ни была наша субстанция — чем-то духовным, материей или швейцарским сыром, она не накладывает никаких существенных ограничений на наш ответ”.

“Я не критикую аналитическую философию за метафизичность, я не борец с метафизикой. Я вижу проблему аналитической философии в том, что она пуста. Философия не может быть одним доказательством, и доказательства не должны быть все в аналитическом стиле. У Кьеркегора, например, есть свои доказательства, хотя аналитические философы никогда этого не признают. Это же верно и в отношении Витгенштейна: его аргументы часто носят педагогический характер, их цель — не объяснить что-либо читателю, а заставить читателя самому сделать выводы. В этом, я думаю, и состоит подлинное назначение философии”.

ХИЛАРИ ПАТНЭМ

ФИЛОСОФИЯ СОЗНАНИЯ

ДОМ
ИНТЕЛЛЕКТУАЛЬНОЙ
КНИГИ

МОСКВА



Данное издание осуществлено в рамках программы Центрального Европейского Университета «Translation Project» при поддержке Регионального издательского центра Института «Открытое общество» (OSI — Budapest) и Института «Открытое общество. Фонд Содействия» (OSIAF — Moscow)

Выражаем благодарность
за информационную помощь в подготовке издания Американскому центру
(г. Москва, ул. Николаямская, д. 1, тел. (095) 956.32.60)

Патнэм Хилари

П 202 **Философия сознания.** Перевод с англ. Макеевой Л. Б., Назаровой О. А., Никифорова А. Л.; предисл. Макеевой Л. Б. — М.: Дом интеллектуальной книги, 1999. — 240 с.

Первый на русском языке сборник работ широко известного современного американского философа смог отразить, к сожалению, лишь часть интересов Х. Патнэма и высказываемых им точек зрения. Однако нам удалось, пожалуй, главное — представить ряд проделанных им и наиболее известных невероятных мыслительных экспериментов, которые уже прочно заняли свое место в современной философии: «Двойник Земли», «машина Тьюринга», «мозги в бочке»...

ББК 87.3

ISBN 5-733-0004-3

© Перевод с англ. коллектив авторов, 1999
© Предисл. Макеевой Л. Б., 1999
© Дом интеллектуальной книги, 1999

СОДЕРЖАНИЕ

Предисловие

7

Сознание и машины

23

Психологические предикаты
(Природа ментальных состояний)

53

Ментальная жизнь некоторых машин

68

Философия и наша ментальная жизнь

88

Философия логики

103

Как нельзя говорить о значении

146

Значение «значения»

164

Предисловие

Многие из тех, кто пишет сегодня об аналитической философии, указывают на симптомы ее кризисного состояния. И хотя внешне аналитическая философия производит впечатление вполне respectable философского направления: она прочно удерживает позиции во многих университетах (по крайней мере, в США и Великобритании), многие известные мыслители продолжают называть себя аналитическими философами, не уменьшается число публикаций, где применяется аналитический подход, и т. д. — тем не менее все чаще слышны высказывания о тупике, в который зашли исследования по той или иной проблеме, об отсутствии хорошо просматриваемой перспективы дальнейшего развития и т. д. Изменился тон произведений аналитических философов — в них стало меньше с энтузиазмом выдвигаемых проектов и больше скепсиса и самокритичности. Настоящий период в развитии аналитической философии настолько не похож на все предыдущие, что для него придумали новое название — «постаналитическая философия». Стоит ли за этим названием та же самая философская парадигма, критически переработанная и по-новому осмысленная, или же мы переживаем период возникновения нового философского направления? Сейчас сложно ответить на этот вопрос. Возможно, многое из того, что происходит в настоящий момент с аналитической философией, станет понятней, если мы обратимся к ее более раннему и более «благополучному» периоду развития — к 60-70-м годам, которые были ознаменованы необыкновенно богатым разнообразием новых идей и теорий — в философии науки и философии языка, в философии сознания и гносеологии. Для нас такой ретроспективный обзор тем более важен, что многие идеи, выдвинутые в тот период, не нашли в нашей отечественной философской литературе достойного внимания и осмысления. Именно поэтому, на наш взгляд, столь своевременно и актуально сейчас знакомство с творчеством известного американского философа наших дней — Хилари Патнэма.

Хилари Патнэм интересен для нас во многих отношениях. Воспитанный в традициях аналитической философии (своими учителями он считает Р. Карнапа и Г. Рейхенбаха), он в течение долгого времени принадлежал к числу ее ведущих представителей. Его работы дают нам яркий образец аналитического стиля философствования. Патнэма отличает высокое мастерство построения изощренных философских аргументов; у него прочная репутация блестящего

оппонента и энергичного участника философских дискуссий. С другой стороны, творческий путь Патнэма сложен и извилист. Со второй половины 80-х годов для Патнэма наступил период критического осмысления задач философского исследования, завершившийся на рубеже 90-х годов декларацией о «смерти» того направления, в авангарде которого он шел целых два десятилетия. С полной уверенностью Патнэм констатировал, что «в тот самый момент, когда аналитическая философия была признана «доминирующим направлением» в мировой философии, она достигла конца своей собственной программы, который оказался не ее выполнением, а «смертельным исходом»¹. Как и почему Патнэм пришел к этому выводу, что побудило его порвать с аналитической философией — ответ на эти вопросы, видимо, поможет лучше понять то «кризисное состояние», которое переживает сейчас доминирующее направление современной философской мысли. Но не менее интересно и полезно было бы обратиться к тому времени, когда была сформирована ее «программа», и рассмотреть, какие цели ставили аналитические философы, на что возлагали надежды, и почему затем наступило разочарование. Работы Патнэма, особенно раннего периода, в этом отношении очень показательны для современного исследователя.

Однако отсюда вовсе не следует, что эти работы Патнэма интересны только с исторической точки зрения. Они представляют собой весомый вклад в разработку разнообразных философских проблем, они обогатили и углубили философское осмысление этих проблем. Без знания идей и аргументов, выдвинутых Патнэмом, нельзя осознать всю сложность и глубину современного уровня философского анализа. Более того, Патнэм обладает качествами, благодаря которым знакомство с его работами становится особенно плодотворным. Как отмечал, известный немецкий исследователь современной философии Вольфганг Штегмюллер, «соединение разнообразных черт, столь удачное, сколь и исключительное, определило центральное место Патнэма в интеллектуальных дискуссиях современного англоязычного мира. Главной среди этих черт является его безошибочное чутье на то, что в необозримом многообразии современных дискуссий имеет подлинную ценность, чутье, которое сочетается у него со способностью в такой манере подойти к проблемам, что это неизменно обещает продвинуть нашу мысль в некото-

¹ *Putnam H. Realism with a Human Face. Cambridge (Mass.), etc.: Harvard Univ. Press, 1990, p. 51.*

проблемам, что это неизменно обещает продвинуть нашу мысль в некотором новом направлении»². Патнэм очень чутко относится ко всему новому, что появляется в философии, но это не означает, что он подстраивается под «преобладающую интеллектуальную программу нашего времени», он один из тех, кто формирует эту программу, постоянно «генерируя» новые решения. Хотя впоследствии Патнэм отказался от многих идей, которые он разрабатывал в своих ранних работах, но эти идеи и эти работы не утратили своего значения и представляют сейчас не только исторический интерес. Знакомство с любыми работами Патнэма стимулирует мысль, открывает новые горизонты, заставляет задуматься над тем, что казалось вполне очевидным.

Перечисленные соображения отчасти объясняют, чем руководствовались составители настоящего сборника, отбирая статьи, опубликованные Патнэмом в период с начала 60-х до середины 70-х годов. Прежде, чем кратко охарактеризовать эти работы и показать, какое место они занимают в творчестве Патнэма, нам хотелось бы дать некоторое общее описание творческого пути этого интересного американского философа³.

Биография Патнэма не богата событиями. Он родился в 1926 году в семье известного переводчика. В 1961—65 гг. работал профессором Массачусетского технологического института, а с 1965 года и по настоящий день является профессором Гарвардского университета. На этом фоне интеллектуальная жизнь Патнэма выглядит несравнимо более насыщенной и интересной. Он автор огромного числа философских трудов; ему принадлежит честь создания ряда философских концепций. Спектр его философских и научных интересов чрезвычайно широк и включает исследования по философским проблемам математики, физики и квантовой механики, исследования по методологии науки и философии языка, разработку проблем логики, философии сознания и этики. Патнэм рассуждает о природе математических истин и защищает реализм

² *Stegmuller W. Hauptstroemungen der Gegenwartsphilosophie. Bd. II, S. 345* (Цит. по: *Putnam H. Realism with a Human Face*, p. xxxix).

³ Более подробное описание творческого пути Патнэма можно найти в книге: *Макеева Л. Б. Философия Х. Патнэма*. М., ИФРАН, 1996. Кроме того, много интересных сведений о его жизни и взглядах по самым разнообразным вопросам можно почерпнуть в книге, которая скоро появится в печати: *Боррадори Дж. Американский философ: Беседы с Куайном, Дэвидсоном, Патнэмом, Рорти, Куном, ...* М.: Дом интеллектуальной книги, 1998.

против номинализма и конвенционализма в математике. Он стремится осмыслить такой важный аспект научного познания как пересмотр научных теорий и пытается согласовать его с научным реализмом. Патнэм критикует бихевиоризм и отстаивает функционализм в решении психофизической проблемы. Он является автором новой концепции референции и страстным критиком традиционной теории значения. Он рассуждает об истине и рациональности, анализирует понятие объективности человеческого знания и критикует метафизический реализм и релятивизм. Он размышляет о природе философских проблем и о судьбах философии в 20-м веке. Он анализирует проблемы философии восприятия и опровергает антиреализм. Это перечисление можно было бы продолжить и дальше, но сказанного достаточно, чтобы признать многогранность и широту творчества Патнэма и понять, насколько хорошо он знает всю «палитру» современных философских изысканий.

Если попытаться выявить во всем этом многообразии основные, «стержневые» темы в творчестве Патнэма, то в их число, безусловно, войдут теория значения, реализм и философия сознания. Именно с этими темами связан главный и наиболее интересный вклад Патнэма в разработку современных философских проблем. Вокруг этих тем и вращается проблематика статей, отобранных для настоящего сборника.

Непосредственно проблеме реализма ⁴ посвящена одна статья в настоящем сборнике — «Философия логики», однако важное значение для понимания позиции Патнэма в этом вопросе имеют

⁴ Как многие аналогичные термины, реализм в философии охватывает целый спектр проблем и связанных с ними позиций. В наиболее общей формулировке он означает признание существующей вне сознания и независимой от него реальности. Однако в зависимости от того, о реальности каких сущностей идет речь и в каком ракурсе рассматривается отношение между сознанием и внешним миром, реализм дробится на множество различных, хотя и связанных друг с другом, проблем. Реализм является проблемой метафизики, если он касается онтологического статуса таких сущностей как универсалии, объекты, постулируемые научными теориями, предметы окружающего нас мира. Проблема реализма обретает гносеологический характер, когда она рассматривается в контексте обоснования возможности достоверного знания. Будучи связанным с вопросом об истинности предложений нашего языка, реализм образует важную тему в философии языка. Если же независимое существование внешних объектов анализируется под углом зрения их «включенности» в наш опыт, то реализм выступает как ключевая проблема философии восприятия и сознания.

его работы «Как нельзя говорить о значении» (1965) и «Значение “значения”» (1975). Это связано с тем, что в целом подход Патнэма к обоснованию реализма можно назвать логико-семантическим, поскольку в нем превалирующее значение имеет анализ референциального значения различных видов языковых выражений, то есть объективное существование разного рода объектов и сущностей обосновывается через призму отношения знака к обозначаемому. Все эти статьи написаны в период, когда Патнэм был убежденным сторонником научного реализма. Однако в последующем Патнэм не только отказался от этой позиции, но и стал одним из наиболее активных и серьезных ее критиков. Впрочем, это не означало его перехода в лагерь противников реализма. Отказ от научного реализма, вызванный тем, что в рамках этой позиции не удалось найти приемлемого решения ряду важных проблем (например, возможности эмпирически эквивалентных, но логически несовместимых теорий и т. д.), ознаменовал начало поиска более адекватной позиции, которая, с одной стороны, сохраняла бы наши реалистические интуиции, а с другой, учитывала бы современный уровень философского осмысления ключевых проблем человеческого бытия и познания. О том, насколько труден этот поиск и насколько последователен Патнэм в своей решимости «провести корабль реализма» между «Сциллой» догматизма и «Харибдой» релятивизма, говорит то многообразие концепций, которые он выдвигал и отстаивал в разные периоды своего творчества: «научный реализм», «внутренний реализм», «реализм с маленькой буквы», «естественный реализм» и т. д. Это многообразие выдвигаемых Патнэмом концепций реализма не следует воспринимать как неумение твердо держаться философских убеждений. Сегодня, как и в начале своего творчества, Патнэм убежден в правильности реализма, но сегодня он, безусловно, иначе понимает и саму проблему реализма, и пути ее решения. Сегодня, как и тогда, Патнэм не сомневается, что правильное решение проблемы реализма лежит не на одном из полюсов «объективное-субъективное», «догматизм-релятивизм», а между ними. Но как проложить этот «средний путь», как вырваться из оков указанных дихотомий — подходы к решению этих вопросов, бесспорно, получают разную трактовку в разные периоды творчества Патнэма. Но вернемся к тому времени, когда Патнэм причислял себя к научным реалистам. Следует отметить, что в этот период реализм был для него не столько темой для размышлений, сколько позицией, в правильности которой он не сомневался. Патнэм видел свою задачу «не в том, чтобы бить себя в грудь по пово-

ду правильности реализма, а в том, чтобы решать конкретные вопросы в философии науки с определенной реалистической позиции»⁶. Однако он подчеркивает, что он реалист, а не материалист, ибо признает не только объективное существование материальных объектов, но и защищает реалистическую позицию в отношении существования математических объектов и физических величин.

«Научный реализм» — термин, вошедший в философский обиход в середине 20 века, однако, не будучи связанным с какой-либо строго определенной и детально разработанной философской доктриной, он выражает общее представление о науке и научных теориях, которое было преобладающим в западной культуре с конца 17 столетия. Согласно этому воззрению, человеку противостоит независимый от него объективный мир; причем этот мир не ограничен только материальными объектами, доступными человеческому восприятию, он имеет «глубины и измерения», скрытые от человеческих чувств и познаваемые наукой. Это означает, что постулируемые истинными научными теориями объекты имеют столь же реальное существование, что и предметы нашего непосредственного восприятия, то есть протоны, фотоны, энергетические поля и черные дыры столь же реальны, как турбины, вулканы, потоки воды и т. д. Более того, эти постулируемые объекты и сущности имеют более фундаментальный характер, поскольку они позволяют объяснить наблюдаемые явления и раскрыть их внутренние механизмы. Хотя указанная сфера реальности выходит за пределы того, что доступно человеческому восприятию, человек может иметь точную и надежную информацию о ней благодаря науке, которая, опираясь на строгие процедуры и методы, раскрывает свойства и характеристики ненаблюдаемых сущностей, тем самым давая истинное знание о них.

Научный реализм возник в западной философии науки как реакция на инструментализм в трактовке научных теорий, а затем стал основным оппонентом в споре с «исторической школой» (Кун Т., Фейерабенд П., Агасси Дж. и др.) о характере развития научного знания. Если научные реалисты придерживались кумулятивной модели развития научного знания, постепенно приближающегося к истинной картине мира, то представители «исторической школы» противопоставили ей представление об истории науки как о процессе смены обособленных концептуальных структур (парадигм,

⁶ Putnam H. *Mathematics, Matter and Method*. Philosophical Papers, Vol. 1, 2d ed. Cambridge, etc.: Camb. Univ. Press, 1975, p. vii.

научно-исследовательских программ, альтернативных теорий), в ходе которого не происходит приращения знания, поскольку знание, накопленное, к примеру, предыдущей парадигмой, отбрасывается в момент ее замены. Патнэм принял активное участие в полемике между представителями исторической школы и научными реалистами. Его статья «Как нельзя говорить о значении» явилась ответной критикой в адрес «исторической школы» и одновременно защитой научного реализма. Его метод защиты состоял в обосновании того, что вывод «исторической школы» (в лице П. Фейерабенда) о несоответствии кумулятивной модели реальной истории науки и о несоизмеримости научных теорий опирается на некорректные посылки, в частности — на неправильную концепцию значения. Но если статья «Как нельзя говорить о значении» в большей мере интересна своими критическими аргументами, то работа «Философия логики» содержит аргументированную защиту тезиса о существовании объектов, постулируемых математическими и физическими теориями (в частности, о существовании чисел, функций и т. п.). Основная идея этого аргумента состоит в том, что без постулирования существования упомянутых абстрактных объектов современная наука невозможна в принципе, то есть невозможно сформулировать большую часть научных положений и законов. Попытки трактовать эти абстрактные сущности как своего рода «полезные фикции» или конвенционально принимаемые объекты не позволяют создать язык, который удовлетворял бы нуждам науки. Статья «Философия логики», несомненно, — блестящий образец аналитического стиля философствования, сочетающего в себе ясность и филигранную точность анализа, глубину философских обобщений и высокое мастерство контраргумента.

Второй наиболее важной темой в творчестве Патнэма являются исследования по проблеме значения. Патнэм входит в число создателей так называемой «новой теории референции», положившей начало «антименталистской» критике традиционной теории значения, которая, можно без преувеличения сказать, является одним из наиболее важных событий в аналитической философии последнего времени.

Параллельно с Патнэмом, хотя и независимо от него, несколько других философов (Доннелан К., Крипке С., Каплан Д. и др.) выдвинули близкие идеи по проблеме значения, что и послужило основанием для объединения этих идей под общим названием «новой теории референции». Несмотря на очевидное концептуальное сходство, подходы упомянутых авторов, безусловно, имели и

различия, обусловленные во многом тем, что они применялись к анализу различных категорий языковых выражений: имен собственных (Крипке, Доннелан), терминов естественных видов (Крипке, Патнэм), индексальных выражений (Каплан).

Мы сможем лучше понять место новой теории референции в контексте современных исследований по проблеме значения, если посмотрим на нее как на проявление одной важной тенденции в развитии философии языка. Со времен Декарта и Локка философы, как правило, придерживались той точки зрения, что слова и выражения нашего языка являются знаками идей и используются прежде всего для выражения наших мыслей. Дж. С. Милль и Г. Фреге положили конец такому пониманию языка, провозгласив, что наши слова служат для обозначения объектов в реальности, а не идей в нашем сознании. Основной функцией языка, с их точки зрения, является вычленение предметов в окружающем нас мире с тем, чтобы высказывать о них истинные утверждения. Поместив на место ментальных образов предметы окружающего мира, Милль и Фреге, однако же, не полностью изгнали идеи из теории значения. Так, согласно Фреге, значение слова представляет собой двухкомпонентное образование: слово обозначает некоторый объект и выражает некоторый смысл (или идею), т. е. то, что мы мысленно схватываем, когда понимаем слово. Это понятие «смысла» (у Милля — понятие «коннотации») и образует тот «менталистский» элемент, который позволил ряду современных философов отнести теорию значения Фреге к картезианской «менталистской» традиции. Этот менталистский элемент еще более усилен постулатом Фреге о том, что смысл языкового выражения определяет его предметное значение или, в современной терминологии, его референцию⁶. Это означает, что референтом слова будет тот объект, который удовлетворяет характеристикам, включенным в смысл этого слова, или, иными словами, удовлетворяет тем дескрипциям, которые связываются с данным словом. Своей семантической концепцией Фреге во многом задал парадигму всех последующих рассуждений о значении в рамках аналитической философии. В то же время в развитии философии языка после Фреге просматривается отчетливая тенденция избавиться от понятия смысла и, таким образом, дове-

⁶ В современной философской литературе вместо фрегевского предметного значения и смысла чаще используются понятия экстенционала и интенционала, введенные Р. Карнапом. Хотя между этими парами понятий есть различия, в рамках обсуждаемой тематики они несущественны.

сти до конца дело, начатое Миллем и Фреге. Это объясняется не в последнюю очередь тем, что понятие «смысл» не поддается выражению в формальном виде и его трудно анализировать логико-математическими методами. Поэтому аналитические философы, для которых строгость и точность анализа всегда были важными атрибутами метода философствования, стремились свести к минимуму или вообще устранить из теории значения понятие смысла. История аналитической философии в 20 веке знает немало попыток представить отношение между языком и миром как прямое, не опосредованное никакими ментальными сущностями⁷. Новая теория референции является очередной попыткой в этом направлении. Ее главный тезис не нов: референция важнейших категорий языковых выражений (имен собственных, терминов естественных видов и индексальных выражений) устанавливается без посредничества смысла. Новизну этой теории составляют способ обоснования этого тезиса и предложенный взамен традиционному (фрегевскому) механизм установления референции.

Наиболее полно семантическая концепция Патнэма изложена в статье «Значение “значения”». Эта семантическая концепция имеет необычную судьбу в творчестве Патнэма, для которого характерен неустанный философский поиск и связанная с ним довольно частая смена взглядов. Меняя свою позицию по тому или иному вопросу, Патнэм не устает подчеркивать приверженность, в главном, своим прежним семантическим идеям. Создается впечатление, что теория референции представляет собой то небольшое в убеждениях Патнэма, что сохраняет «инвариантность» на всем протяжении его богатой изменениями творческой биографии. Но это не совсем так. Поскольку эта теория занимает важное место в обосновании предложенных им в разное время концепций реализма, при всей неизменности формулировок ряда ключевых положений этой теории, они получают очень разное истолкование в контексте этих концепций реализма. Более того, в ходе философской эволюции Патнэма усиливается антименталистская тональность его семантической концепции.

Однако в период написания работы «Значение “значения”» антиментализм Патнэма выражался лишь в несогласии с традиционной трактовкой механизма установления референции слов. Патнэм убежден, что значения слов, представленные «ментальными образами» в сознании человека, не могут определять референцию этих

⁷ Достаточно упомянуть здесь Рассела и Витгенштейна.

слов. Но он еще не готов полностью отказаться от «смыслового» компонента в значении слов и поэтому выбирает стратегию дальнейшей дифференциации понятия значения. Признавая, что исчерпывающее определение значения — это задача будущего научного исследования, Патнэм предлагает лишь его общую идею — представить значение как «вектор», образованный из множества компонентов: синтаксических и семантических маркеров, стереотипа, дескрипции экстенционала и т. д. Примечательно, что стереотип почти ничем не отличается от фрегевского смысла за тем исключением, что он не определяет референцию слов.

Отказавшись от смысла как механизма, определяющего и систематически обеспечивающего референцию различных видов языковых выражений, сторонники новой теории референции оказались перед необходимостью предложить иной механизм. Обобщенно их основной тезис в решении этой проблемы можно сформулировать так: референция указанных выражений устанавливается благодаря внешним нементальным факторам.

Так, согласно Патнэму, в установлении референции терминов естественных видов участвуют два фактора: социальный (в силу того, что существует «разделение лингвистического труда») и природный (благодаря тому, что «сами естественные виды играют определенную роль в установлении экстенционалов терминов, которые на них указывают»⁸).

Действие социального фактора Патнэм описывает с помощью «социолингвистической гипотезы». Согласно этой гипотезе различные категории носителей языка знают и используют разные аспекты значения слов. Поэтому референцию слов определяют не обрывочные и неполные «значения», локализованные в головах отдельных людей, а та детальная информация и методы распознавания, которыми владеют эксперты.

Действие второго «природного» фактора в определении референции слов (по крайней мере, терминов, обозначающих естественные виды: природные вещества, растения, физические величины и животных) связано с тем, что любой естественный вид предполагает наличие у его членов общей внутренней природы (или сущности), выражающейся в общей внутренней структуре, общих естественных свойствах или общих объективных законах, управляющих поведением или развитием членов данного естественного вида. Это

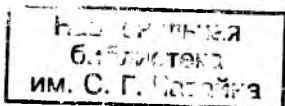
⁸ *Putnam H. Reference and Truth // Putnam H. Realism and Reason. Philosophical Papers, Vol. 3. Cambridge: Camb. Univ. Press, 1983, p. 74.*

означает, что в экстенционал термина естественного вида входят те объекты, которые обладают внутренней природой, присущей данному естественному виду, то есть естественные виды «сами выполняют работу» по установлению экстенсионалов своих терминов. Этот природный фактор предполагает каузальный механизм установления связи между носителями языка и референтами используемых ими слов. Вначале всегда имеет место каузальное взаимодействие с определенными представителями естественного вида, которым присваивается некоторый термин и которые таким образом становятся парадигмальным образцом этого вида и по отношению к которым устанавливается принадлежность других объектов к этому виду.

Этот аспект новой теории референции дал повод для многочисленных обвинений Патнэма (и Крипке, который придерживался сходной позиции) в эссенциализме. Признание внутренней природы, конституирующей естественный вид, было расценено как возврат к «давно дискредитировавшей себя» локковской (или даже аристотелевской) идее «реальной сущности».

В дальнейшем, отчасти под влиянием этой критики, отчасти в ходе напряженных раздумий над проблемами референции, реализма и истины Патнэм изменил трактовку основных положений своей семантической концепции. В книге «Разум, истина и история» («Reason, Truth and History», 1981) он отказывается от определения значения как четырехкомпонентного образования, содержащего такой элемент, как стереотип, который имеет прямое отношение к сфере ментального. Он предлагает иную трактовку тезиса о том, что значения не находятся в голове говорящих, — трактовку более близкую по духу идеям Витгенштейна. Согласно этой трактовке, понятия, к которым значение наших слов имеют прямое отношение, представляют собой знаки, употребляемые ситуативно надлежащим образом, а вовсе не «ментальные репрезентации, которые с необходимостью указывают на внешние объекты»⁹. Получает совсем иное истолкование и тезис о том, что природа сама устанавливает референцию наших терминов. Во избежание обвинений в эссенциализме Патнэм указывает, что «расчленение» мира на объекты разных видов определяется нашими концептуальными схемами. Экстенционал любого термина содержит не только объекты, с которыми мы имели каузальное взаимодействие, но и все объекты

⁹ Putnam H. Reason, Truth and History. Cambridge, etc.: Camb. Univ. Press, 1981, p. 18.



того же самого вида. Однако выражение «того же самого вида» утрачивает смысл вне категориальной системы.

Однако и это решение в дальнейшем будет признано Патнэмом неудовлетворительным в силу того, что в нем еще очень силен менталистский элемент. В начале 90-х гг. Патнэм¹⁰ признает, что своей прежней позицией ему не удалось полностью преодолеть картезианский взгляд на природу ментального как на некоторую область взаимодействия между нашими когнитивными способностями и объектами внешнего мира. Этот картезианский подход, согласно Патнэму, не оставляет иной возможности объяснить референцию языковых выражений как каузальным взаимодействием человека с внешними объектами, обозначенными этими выражениями. Однако учитывая, что значения истолковываются здесь как «ментальные образы», локализованные исключительно внутри сознания, каузальное объяснение референции оказывалось задачей столь безнадежной, что оно не раз толкало философов к принятию того или иного варианта идеализма — в частности толкнуло его, Патнэма, к признанию зависимости внешнего мира от сознания, от наших концептуальных схем и т. д. Новое решение, предлагаемое Патнэмом, уже полностью в духе Витгенштейна.

Согласно этому решению, значение лингвистических выражений является тем, что *показывает себя* в наших словах и предложениях. Когда мы слышим или читаем слова и предложения, мы не воспринимаем их как простые «звуки и знаки», в которые должно быть «вставлено» значение, находящееся вне их и присутствующее в нашем сознании как некая «ментальная сущность». Мы воспринимаем значение в самих словах и предложениях, но отсюда не следует, считает Патнэм, что значение присуще им от природы. Наши слова и предложения обладают значением, потому что имеет место определенная «техника употребления», благодаря которой значение *показывает свое лицо* в них. Здесь Патнэм использует идею Витгенштейна о том, что мы можем «видеть лицо» одной деятельности в другой (как мы *видим* изображение человеческого лица в различных конфигурациях линий и точек). Одна деятельность может *показывать себя* в другой благодаря тому, что все виды человеческой деятельности тесно увязаны друг с другом, образуя сложную и разветвленную систему. Поэтому и мыш-

¹⁰ См.: Putnam H. Sense, Nonsense, and the Senses: An Inquiry into Powers of the Human Mind // Journal of Philosophy. 1994, Vol. 91, № 9, p. 445—517.

ление — это не отдельно стоящая деятельность, не поддерживаемая никакими другими видами деятельности. Оно вплетено в сложную систему практик как лингвистических, так и нелингвистических.

Итак, эволюция семантических идей Патнэма отражает стремление избавиться теорию значения от понятия смысла и связанных с ним менталистских допущений. Мы не будем здесь оценивать, насколько успешным оказалось это очередное наступление против традиционной теории значения. Отметим лишь один момент: достигнутое в результате решение, на деле, оказывается простым устарением проблемы значения: мы не можем объяснить «менталистский» аспект значения — следовательно, этого менталистского аспекта вообще нет. Безусловно, это очень радикальное решение; насколько оно адекватно — в этом еще предстоит разобраться.

Третьей важнейшей теме в творчестве Патнэма — философии сознания — посвящены включенные в настоящий сборник статьи «Сознание и машины» (1960), «Ментальная жизнь некоторых машин» (1967), «Психологические предикаты» (1967) и «Философия и наша ментальная жизнь» (1973). Эти статьи позволят читателю войти в круг проблем, обсуждаемых в современной аналитической философии сознания. К сожалению, эти проблемы нашли крайне скудное освещение в отечественной философской литературе, и это создало дополнительные трудности при переводе указанных статей, ибо у нас нет пока устоявшейся терминологии в этой области. Поэтому целесообразно сказать несколько слов о переводе наиболее важных терминов.

На наш взгляд, наибольшую трудность составляет такой английский термин как «*mind-body problem*». При переводе этого термина мы руководствовались следующими соображениями. Во-первых, в современной аналитической философии термин «*mind-body problem*» соотносится не с одной конкретной проблемой, а с некоторым, так сказать, проблемным «полем», которое у разных философов получает очень разные истолкования. Поэтому в качестве перевода нужен термин достаточно широкий и неопределенный, который не навязывал бы единственной жесткой интерпретации. Во-вторых, не следует забывать, что многие авторы подчеркивают традиционный характер «*mind-body problem*». На наш взгляд, в русском философском «словаре» лучше всего этим требованиям отвечает термин «психофизическая проблема». Это термин связан с классической философией и достаточно неопределенен, чтобы, очерчивая соответствующее проблемное поле, не навязывать жест-

кой интерпретации. Что же касается английского термина «*mind-body identity*», который имеет отношение к довольно недавней «теории тождества» (*identity theory*), то его содержание лучше всего передает русский термин «тождество сознания и мозга». Дело в том, что в указанной теории, как правило, говорится о тождестве состояний сознания и состояний мозга (*brain*), а если же подчеркивается, что речь идет о состояниях всего организма, то они чаще всего трактуются как состояния нервной системы. Поэтому хотя «мозг» — более узкий термин, чем «нервная система», для характеристики указанного тождества термин «мозг» более уместен, чем термин «тело», имеющий множество нежелательных коннотаций.

Психофизическая проблема составляла важный предмет исследования на всем протяжении творческого пути Патнэма. В начале 60-х годов он предложил, параллельно с несколькими другими философами, новый подход к решению этой проблемы, который получил название «функционализма». В течение длительного времени Патнэм отстаивал эту позицию, однако в 80-е годы он уже основное внимание уделял ее критическому анализу, завершившемуся в начале 90-х годов полным отказом от нее.

Функционализм представляет собой вариант уже упоминаемой теории тождества (*identity theory*), которая в настоящее время является преобладающей методологией в философии сознания и когнитивной науке и согласно которой наши ощущения и восприятия тождественны определенным состояниям мозга. В отличие от физикалистского варианта теории тождества функционализм отождествляет ментальные состояния не с физическими состояниями мозга, а с его «функциональными» состояниями. При таком подходе мозг понимается как очень сложная система, которая помимо физических свойств обладает свойствами более высокого уровня — так называемыми функциональными свойствами. Функционализм предполагает прямую аналогию между мозгом и «цифровым компьютером»: «наша психология должна быть описана как программное обеспечение этого компьютера — как его «функциональная организация»¹¹. Главная особенность функционализма состоит в том, что он «определяет ментальные состояния и процессы через их каузальные и функциональные отношения друг к другу, к перцепту-

¹¹ Putnam H. *Representation and Reality*. Cambridge (Mass.): MIT Press, 1989, p. 73.

альным «входным данным», поступающим из внешнего мира, и к поведенческим «выходным данным», выражаемым в действии»¹².

Предлагаемые вниманию читателя статьи Патнэма по философии сознания отражают определенную эволюцию его взглядов. В статье «Сознание и машины» где Патнэм, по сути, дает первую формулировку функционального подхода к проблеме сознания, он предлагает использовать для описания функциональных состояний мозга формализмы теории машин Тьюринга. Машину Тьюринга можно охарактеризовать как множество команд для выполнения множества простых операций над цепочками символов, которые образуют «входные данные» машины. Эти команды группируются в «машинные состояния», каждое из которых, будучи конечным множеством, управляется некоторой главной командой, переключающей состояния и определяющей их последовательность в соответствии с «входными данными». Важно подчеркнуть, что одна и та же машина Тьюринга может быть реализована в разных физических системах (изготовленных из разных материалов, с разными структурными свойствами и т. д.).

Однако очень скоро обнаружилось, что формальные свойства таким образом описанных «машинных состояний» сильно отличаются от формальных свойств ментальных состояний. Патнэм попытался дать новую формулировку функционализма в виде «психологической теории», которая опиралась на использование формальных средств теории вычислений. Этот сдвиг во взглядах Патнэма отражает статья «Философия и наша ментальная жизнь». Но как было установлено впоследствии, этот вариант функционализма тоже оказался неудовлетворительным¹³.

Неудачи в реализации функционального подхода к проблеме сознания заставили Патнэма задуматься над теми общими допущениями, которые лежали в основе этого подхода, и поставил перед

¹² *Harman G. The Intrinsic Quality of Experience // Philosophical Perspectives. 1990, Vol. 4, p. 32.*

¹³ Теория вычислений позволяет имплицитно определить и индивидуализировать любое состояние системы через всю совокупность его вычислимых отношений ко всем другим состояниям системы, однако никакая существующая психологическая теория не позволяет сформулировать совокупность законов, которые бы отличали, скажем, состояние ревности какого-то человека от любого другого психологического состояния. Поэтому установление тождества между состояниями, описываемыми теорией вычислений, и психологическими состояниями является безнадежной задачей.

ним задачу поиска иной «парадигмы» для исследования отношения между сознанием и мозгом. Эти поиски, вплетенные в контекст его размышлений о реализме, изменили как его представление о самом реализме, так и понимание им проблемы сознания.

Сегодня главное препятствие на пути решения психофизической проблемы Патнэм видит в господствующем картезианском представлении о сознании. Согласно Патнэму, существо картезианского взгляда на природу ментального составляет трактовка нашего опыта как принадлежащего исключительно сфере сознания, сфере, которая находится внутри нас и которая имеет нефизическую природу. Эта трактовка подкрепляется каузальной теорией восприятия, согласно которой объекты внешнего мира, воздействуя на наши органы чувств, причинно вызывают цепочку событий, ведущую к возникновению «ментальных образов» в нашем сознании.

Главное обвинение Патнэма направлено против неспособности картезианской философии сознания, в течение трех веков «кутающейся в мантию науки», предложить, помимо пустого разговора о «концептуальной структуре сознания», хоть какое-нибудь маломальски приемлемое решение для тех проблем, которые ею же и были сформулированы. Согласно Патнэму, теория тождества не может предложить никакого механизма, объясняющего порождение ментальных репрезентаций событиями или процессами в мозге. Не способна она объяснить и то, как мы непосредственно наблюдаем и осознаем постулируемые ею ментальные сущности.

Патнэм не одинок в своей характеристике современного состояния философии сознания и когнитивной науки как «тупикового». Многие осознают, что существующая методология решения проблемы сознания уже исчерпала свои возможности, так и не предложив адекватных решений, поэтому необходим принципиально новый взгляд на отношения между сознанием и мозгом.

Итак, мы кратко описали основные линии в философской эволюции Патнэма. Хотя сегодня Патнэм отказался от многих позиций, которые он отстаивал в статьях, включенных в настоящий сборник, однако эти позиции не утратили своего значения и актуальности в современных философских дискуссиях. Поэтому знакомство с ними окажется плодотворным для всех, кто интересуется проблемами реализма, значения и сознания.

*Макеева Л. Б.,
кандидат философских наук*

Разнообразные вопросы и загадки, составляющие классическую психофизическую проблему (*mind-body problem*), носят исключительно лингвистический и логический характер: если и есть в этой области немногочисленные эмпирические «факты», то они в равной мере подтверждают самые разные точки зрения. Я не ставлю себе цели доказать это утверждение в настоящей статье, но надеюсь сделать его более правдоподобным. В частности, я попытаюсь показать, что все эти проблемы возникают в связи с любой компьютерной системой, способной отвечать на вопросы относительно своей собственной структуры, и никак не связаны с уникальной природой человеческого субъективного опыта (если она действительно уникальна).

В качестве иллюстрации сошлюсь на проблему *приватности* (*privacy*), которая представляет собой одну из загадок, время от времени обсуждаемых в связи с «психофизической проблемой». Вопрос «Откуда я знаю, что мне больно?» является *девиантным*¹ («логически странным») вопросом. Вопрос же «Откуда я знаю, что Смиту больно?» совершенно не девиантен. Это различие можно проследить и в случае безличных вопросов: вопрос «Откуда некто знает, что ему больно?» девиантен, а вопрос «Откуда некто знает, что кому-то другому больно?» не девиантен. Далее я покажу, что различие между двумя последними вопросами воспроизводится и в случае машин: если *T* — машина Тьюринга, то вопрос «Как машина *T* устанавливает, что она находится в состоянии *A*?», как мы увидим в дальнейшем, в еще большей степени «логически странный»; но если машина *T* способна обследовать свою соседку, машину *T'* (скажем, *T* обладает электронными «органами чувств», «сканирующими» машину *T'*), то вопрос «Как машина *T* устанавливает, что *T'* находится в состоянии *A*?» вовсе не является странным.

Другой вопрос, связанный с «психофизической проблемой», формулируется так: допустимо ли вообще отождествлять ментальные и физические события. Конечно, я не утверждаю, что этот во-

* Впервые опубликовано в кн.: *Hook S. (ed.) Dimensions of Mind*, N. Y., 1960.

¹ Девиантным здесь называется предложение, семантически не вполне корректное (в соответствующем естественном языке). Этот термин взят у Зиффа (см. *Ziff P. Semantic Analysis*, N. Y., 1960).

прос правомерен и в отношении машин Тьюринга, но считаю, что применительно к ним можно сформулировать логический аналог вопроса о «тождестве сознания и мозга», который воспроизведет все аспекты этого вопроса.

В целях построения такого аналога определим научную теорию как «частично-интерпретированное исчисление» в смысле Карнапа². Теперь мы вполне можем представить себе машину Тьюринга, которая создает теории, проверяет их (предположим, что в определенной степени можно сформулировать «машинный» вариант индуктивной логики) и «принимает» теории, удовлетворяющие определенным критериям (например, критерию успешности предсказаний). В частности, если машина обладает электронными «органами чувств», позволяющими ей «сканировать» себя во время работы, она может формулировать теории о своей собственной структуре и подвергать их проверке. Предположим, машина находится в каком-то определенном состоянии (скажем, «состоянии А») тогда и только тогда, когда взведен триггер 36. Тогда предложение: «Я нахожусь в состоянии А тогда и только тогда, когда взведен триггер 36» может выражать один из теоретических принципов, принятых машиной для описаний своей собственной структуры. Безусловно, с точки зрения машины, выражение «Я нахожусь в состоянии А» относится к «языку наблюдения», тогда как «взведен триггер 36» представляет собой «теоретическое выражение», частично интерпретируемое в терминах «наблюдаемых сущностей» (если сообщения от «органов чувств» машины печатаются в виде символов на ленте входных данных, то «наблюдаемые сущности», в терминах которых машина формулировала бы операциональное определение для выражения «взведен триггер 36», задающее его частичную интерпретацию, имели бы форму «символ ≠ так-то и так-то появляется на ленте входных данных»). Теперь для всех обычных соображений «за» и «против» отождествления сознания и мозга можно сформулировать сходные соображения «за» и «против» *отождествления* состояния А и взведенного состояния триггера 36.

Параллельно аргументам в духе Оккама, выдвигаемым в защиту «тождества» сознания и мозга, можно предложить сходные, в

² См.: Carnap R. Interpretation of Physics, 1953; The Methodological character of theoretical concepts // Minnesota Studies in the Philosophy of Science, Vol. I, Minneapolis, 1956, p. 38—76. На мой взгляд, эта модель научной теории слишком проста, чтобы иметь широкое применение: однако эта упрощенность не влияет на наше обоснование.

духе Оккама, аргументы в случае машин. И для обычного аргумента в пользу дуализма можно сформулировать следующий параллельный аргумент: с точки зрения машины «состояние *A*» непосредственно наблюдаемо; с другой стороны, знание о «триггерах» она получает только в результате чрезвычайно сложных выводов — Как могли бы две столь различные вещи *быть* тождественными? Этот последний аргумент будет более убедительным, если его сформулировать иначе. Высказывание:

(1) Я нахожусь в состоянии *A*, если и только если взведен триггер 36,

для машины определенно является «синтетическим». Например, машина могла бы находиться в состоянии *A*, а ее органы чувств давали бы сообщение, что триггер 36 *не* взведен. В таком случае машина должна была бы сделать методологический «выбор»: или отказаться от высказывания (1), или заключить, что имела место «ошибка наблюдения» (и человек-ученый, изучая свои собственные психофизические корреляции, оказался бы перед сходным методологическим выбором). Точно так же философы, исходя из синтетической природы высказывания:

(2) Мне больно, если и только если возбуждены мои *C*-волокна,

сделали бы вывод о том, что *свойства* («состояния» или «события») «испытывать боль» и «иметь *C*-волокна возбужденными», видимо, не могут быть тождественными (в противном случае высказывание (2) было бы аналитическим) — примерно так формулируется этот аргумент. Поэтому на том основании, что высказывание (1) является синтетическим, следовало бы заключить, что два свойства («состояния» или «события») — «находиться в состоянии *A*» и «иметь триггер 36 взведенным», — не могут быть тождественными!

Нелишне отметить, что традиционный аргумент в пользу дуализма — это вовсе не вывод на основе «необработанных данных непосредственного опыта» (о чем говорит тот факт, что он с таким же успехом применим к не наделенным сознанием машинам), а довольно сложное рассуждение, которое предполагает: (а) опредме-

чивание (reification) универсалий ³ (например, «свойств», «состояний», «событий») и (б) проведение четкого различия между аналитическим и синтетическим.

Из-за предложенной мной аналогии меня могут обвинить в защите «механистического» взгляда на мир. Если при этом мне будут приписывать ту точку зрения, что машины мыслят ⁴, или что люди представляют собой машины, то это ложное обвинение. Если же имеется вариант механицизма достаточно изошренный, чтобы исключить подобные ошибки, то скорее всего соображения, высказанные в настоящей статье, подтверждают его ⁵.

1. МАШИНЫ ТЬЮРИНГА

В настоящей статье нам потребуется понятие *машины Тьюринга* ⁶, которое я и хотел бы сейчас разъяснить.

В двух словах, машина Тьюринга — это устройство с конечным числом внутренних конфигураций, каждая из которых характеризуется тем, что машина находится в каком-то одном состоянии

³ Этот момент был отмечен Куайном в работе: *Quine W. V. O. The Scope and Language of Science // British Journal for the Philosophy of Science, 1957, Vol. VIII, p. 1—17.*

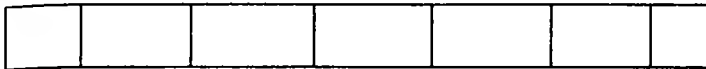
⁴ См. статью Зиффа (*Ziff P. The Feelings of Robots // Analysis, 1959, Vol. XIX, p. 64—68*) и ответ на нее Смарта (*Smart J. Professor Ziff on Robots // Analysis, 1959, Vol. XIX, p. 117—118*). Зифф говорил мне, что под «роботом» он не имел в виду «обучаемую машину» описанного Смартом вида, и он согласен, что выдвинутые в его статье соображения необязательно применимы к подобной машине (если она вообще может быть определена как «машина»). Что касается *девиантности* вопроса «может ли эта машина мыслить (чувствовать и т. п.)?», то здесь необходимо учитывать соображение, высказанное Зиффом (важно не то, девиантно ли предложение или нет, а то, девиантно ли оно по нетривиальным причинам), равно как и различие между «диахроническим и синхроническим», обсуждаемое в разделе 5 настоящей статьи.

⁵ В частности, мне нравится в целом точка зрения, выдвинутая Смартом в его работах: *Smart J. Incompatible Colors // Philosophical Studies, 1959, Vol. X, p. 39—42; Smart J. Sensations and Brain Processes // Philosophical Review, 1959, Vol. LXVIII, p. 141—156.* Тем не менее, см. лингвистические соображения в разделе 5 настоящей статьи.

⁶ Что касается деталей ср. работы: *Davis M. Computability and Unsolvability. N. Y., 1958; Kleene S. Introduction to Metamathematics, N. Y., 1952.*

из конечного множества состояний ⁷ и сканирует ленту, на которой появляются определенные символы.

Лента машины разделена на отдельные клетки следующим образом:



В каждой из этих клеток может быть напечатан символ (из заданного конечного алфавита). Кроме того, машина имеет «сканер», который «сканирует» одну клетку ленты за раз. И, наконец, машина имеет печатающий механизм, который может: (а) удалить символ, встретившийся в сканируемой клетке, и (б) напечатать вместо него какой-то другой символ (из машинного алфавита).

Любая машина Тьюринга полностью описывается машинной таблицей, которая строится следующим образом: строки таблицы обозначены буквами алфавита (включая и «пустую» букву, т. е. пробел), а столбцы соответствуют состояниям A, B, C и т. д. В каждой клетке появляется «команда», например « $s_5L A$ », « $s_7C B$ », « $s_3R C$ ». Эти команды читаются следующим образом: « $s_5L A$ » — «напечатать символ s_5 в клетке, которая сейчас сканируется (после удаления уже содержащегося в нем символа), затем просканировать ближайшую клетку слева от только что просканированной; после этого перейти в состояние A ». Другие команды интерпретируются сходным образом (« R » означает «сканировать ближайшую клетку справа», тогда как « C » означает «центрировать», т. е. продолжать сканировать ту же самую клетку). Примером машинной таблицы может быть следующий:

		A	B	C	D
(s_1)	1	s_1RA	s_1LB	s_3LD	s_1CD
(s_2)	+	s_1LB	s_2CD	s_2LD	s_2CD
(s_3)	пробел	s_3CD	s_3RC	s_3LD	s_3CD

⁷ Эта терминология взята из работы Клини (1952) и отличается от терминологии Дэвиса и Тьюринга.

Предполагается, что машина, описываемая этой таблицей, функционирует следующим образом: при запуске она находится в состоянии A . На ленте появляется «сумма» (в унарном представлении), которую нужно «вычислить», например « $II + III$ ».

Машина сначала сканирует первую цифру « I » и начинает «вычислять» сумму (по сути, заменяя знак плюс на цифру I , а затем возвращаясь и стирая первую цифру I). Так, если бы «входные данные» были $IIII + IIII$, то машина «напечатала» бы $IIIIIIII$, а затем перешла бы в «неактивное состояние» (состояние D).

«Машинная таблица» описывает машину, если та обладает внутренними состояниями, соответствующими столбцам таблицы, и «выполняет» команды из таблицы в том смысле, что сканируя клетку, в которой встречается символ s_1 , и находясь, скажем, в состоянии B , она выполняет «команду», указанную в соответствующей строчке и столбце таблицы (в данном случае, в столбце B и строчке s_1). Любая машина, описываемая машинной таблицей указанного типа, является машиной Тьюринга.

Понятие машины Тьюринга можно расширить в разных отношениях⁸. Например, можно предположить, что у машины есть вторая лента («входная лента»), на которой оператор в ходе вычислений печатает дополнительную информацию. В дальнейшем мы воспользуемся этим обобщением (но место «оператора» у нас займут электронные «органы чувств»).

Примечательно, что машины Тьюринга, в принципе, способны делать все то, что делает любая вычислительная машина (любого вида)⁹.

Иногда утверждают (как, например, Нагель и Ньюман в их книге «Доказательство Гёделя» («*Gödel's Proof*»)), что «эта теорема [т. е. теорема Гёделя] свидетельствует о том, что человеческое мышление обладает гораздо более сложной и совершенной структурой и гораздо большей мощностью, чем любая неживая машина, какую мы могли бы себе представить», и, следовательно, машина Тьюринга не может служить моделью человеческого мышления, но это совершенно ошибочное мнение.

⁸ Такое расширение выполнено Дэвисом в работе: *Davis M. Computability and Unsolvability*. N. Y., 1958, где оно используется для определения относительной рекурсивности.

⁹ Это утверждение представляет собой разновидность тезиса Чёрча (о том, что рекурсивность равнозначна эффективной вычислимости).

Пусть T — машина Тьюринга, которая «репрезентирует» меня в том смысле, что T может доказывать те же самые математические положения, что и я. Тогда (хотя Нагель и Ньюман не приводят никакого доказательства, но, думаю, именно этот аргумент они, должно быть, имели в виду) методом Гёделя можно найти высказывание, которое машина T не может доказать, а я могу доказать. Это опровергает то допущение, что T «репрезентирует» меня: следовательно, я не машина Тьюринга. Ошибка заключается в неправильном применении простой и ясной теоремы Гёделя. Если взять произвольную машину T , то я могу только одно — найти такое высказывание U , для которого я могу доказать:

(3) Если T непротиворечива, то U истинно,

где U неразрешимо для T , если T на самом деле непротиворечива. Однако T вполне может доказать утверждение (3)! А утверждение U , которое T не может доказать (при условии ее непротиворечивости), я также не могу доказать (если только я не смогу доказать, что T непротиворечива, а это маловероятно, учитывая крайнюю сложность T)!

2. ПРИВАТНОСТЬ

Предположим, что машина Тьюринга T сконструирована для выполнения следующих действий: Какое-то число, скажем, «3000» печатается на ленте машины T , и T оказывается в своем «начальном состоянии». Вслед за этим T вычисляет трехтысячную (или любую другую, как указывает заданное число) цифру в десятичном разложении числа π , печатает эту цифру на своей ленте и переходит в «неактивное состояние» (т. е. сама себя выключает). Разумеется, вопрос: «Как T «устанавливает» [«вычисляет» или «определяет»] трехтысячную цифру в десятичном разложении числа π ? — вполне разумен, а ответ на него может быть довольно сложным. Вполне вероятно, ответ будет включать три различных компонента:

(i) Описание последовательности состояний, через которые проходит T для нахождения решения, и описание того, что появляется на ленте на каждом этапе вычислений.

(ii) Описание *правил*, по которым действует T (они задаются «машинной таблицей» для T).

(iii) Объяснение *рациональной основы* всей процедуры.

Предположим, кто-то высказывает следующее возражение: «Для выполнения только что описанных вычислений, машина T должна последовательно находиться в состояниях A , B , C и т. д. Но как T может установить, что она находится в состояниях A , B , C и т. д.?»

Ясно, что это возражение нелепо. Но в силу чего оно нелепо? Во-первых, «логическое описание» машины (ее машинная таблица) определяет состояния только в терминах их *отношений* друг к другу и к тому, что появляется на ленте. «Физическая реализация» машины нематериальна, поскольку *есть* отдельные состояния A , B , C и т. д., и они следуют друг за другом, как указано в машинной таблице. Поэтому, ответить на вопрос: «Как T устанавливает, что X ? (или «вычисляет X » и т. п.)» — можно, лишь описав *последовательность состояний*, через которую должна пройти T , чтобы установить, что X (вычислить X и т. п.), и описав правила, которым она подчиняется, и т. д. Но нет «последовательности состояний», через которую должна пройти T , чтобы оказаться в каком-то отдельном состоянии!

Однако, предположим, что такая последовательность имеется, — предположим, что T не могла бы быть в состоянии A , не установив прежде, что она находится в состоянии A (пройдя через последовательность других состояний). Ясно, что это ведет к регрессу в бесконечность. Мы можем «прервать» этот регресс, просто указав, что машина, устанавливая трехтысячную цифру в числе π , *последовательно находится в определенных состояниях*; но ей не нужно ни в каком существенном смысле «устанавливать», что она последовательно находится в этих состояниях.

Отметим аналогичную ошибку в классической эпистемологии, где предполагается, что для того, чтобы знать, что p (где p — любое высказывание), нужно сначала знать, что q_1 , q_2 и т. д. (где q_1 , q_2 и т. д. — соответствующие случаю *другие* высказывания). Все это ведет к «регрессу в бесконечность» или к изобретению сомнительного особого класса «протокольных» предложений.

Устраняется эта ошибка аналогично тому, как устраняется ошибка в случае с машинами. Предположим, что на основе данных чувственного опыта E_1 , E_2 и т. д., я знаю, что в комнате есть стул. Отсюда не следует ни того, что я вербализую (или даже *мог бы* вербализовать) E_1 , E_2 и т. д., ни того, что я помню E_1 , E_2 и т. д., ни того, что я «мысленно классифицировал» («обратил внимание на» и т. п.) данные чувственного опыта E_1 , E_2 и т. д., когда их ощущал. Короче говоря, для определенного рода знания необходи-

мо *иметь* чувственный опыт, но вовсе не нужно *знать* (или даже *замечать*), какой именно чувственный опыт ты имеешь.

Попробуем, однако, несколько изменить наш пример. Предположим, что, находясь в определенном состоянии (скажем, «состоянии А»), машина печатает слова «Я нахожусь в состоянии А». Тогда, признавая, что хотя машина, как правило, не знает, в каком состоянии она находится, но в случае состояния А (после того, как она напечатала «Я нахожусь в состоянии А») можно сказать: «Машина установила, что она находится в состоянии А».

Рассмотрим данный пример повнимательнее. Прежде всего нам следует предположить, что, находясь в состоянии А, машина печатает «Я нахожусь в состоянии А» не потому, что перед этим она прошла через серию каких-либо других состояний. Просто во всех строчках столбца таблицы, озаглавленного «состояние А», стоит команда: *напечатать*¹⁰ «Я нахожусь в состоянии А». Во-вторых, для сравнения рассмотрим какого-нибудь человека, скажем, Джона, который каждый раз, когда ощущает боль, говорит: «Мне больно» (или «Ох!», или «Больно!»). Ради пущего сходства, предположим, что в лингвистическом плане слова Джона объясняются тем, что он просто «не думая» говорит «Мне больно», т. е. он не проходит через серию каких-либо интроспективных ментальных состояний, помимо состояния самой боли. Используя терминологию Витгенштейна, Джон просто *показывает* свою боль, когда говорит «Мне больно», — перед этим он не рефлексировал о своем состоянии (обращая на него внимание, отмечая его и т. п.), с тем, чтобы затем сознательно его описать. (Заметьте, что эта простая возможность произнесения «высказывания» «Мне больно» без предварительного ментального «акта суждения» осталась незамеченной для всех классических эпистемологов от Юма до Рассела!) Теперь мы можем рассмотреть параллельные вопросы: «"Устанавливает" ли машина, что она находится в состоянии А?» и «"Знает" ли Джон, что он ощущает боль?» и их следствия.

Философы, занимающиеся семантикой, уделяют, как и следовало ожидать, большое внимание глаголу «знать». Традиционно они выделяют три элемента: (1) «Х знает, что р» означает, что р *истинно* (мы можем назвать это элементом *истинности*); (2) «Х знает, что р» означает, что Х верит, что р (слово «верит» вызывает у философов споры. Некоторые из них считают, что следует гово-

¹⁰ Здесь следует принять, что все предложение «Я нахожусь в состоянии А» в алфавите машины считается одним символом.

речь: «*X* уверен, что *p*» или «*X* в состоянии утверждать, что *p*»; я буду называть это элементом *уверенности*); (3) «*X* знает, что *p*» означает, что у *X* есть подтверждение (evidence) того, что *p* (думаю, что здесь слово «подтверждение» безусловно ошибочно¹¹, но для наших целей это не существенно; я буду называть это элементом *подтверждения*). Более того, значение слова «подтверждение» предполагает то, что нельзя быть в буквальном смысле подтверждением для самого себя: если *X* — подтверждение для *Y*, то *X* и *Y* должны быть разными вещами.

В свете такого анализа возникли споры о том, правомерно ли говорить (в случаях, подобных рассмотренному выше), что «Джон знает, что ощущает боль». С одной стороны, философы, стоящие на точке зрения здравого смысла («Когда мне больно, я *знаю*, что мне больно»), рассуждают примерно так: несомненно ложным было бы утверждение, что Джон *не* знает, что ему больно; но, Джон или знает это, или не знает; следовательно, Джон знает, что ему больно. Этим философам можно было бы возразить следующее: из утверждения «Джон не знает *X*» следует, что Джон не в состоянии утверждать, что *X*; поэтому совершенно неверно утверждать, что «Джон не знает, что ему больно». Однако применение закона исключенного третьего было бы в данном случае ошибочным: слова в английском языке имеют *диапазоны значения* (significance gaps), и выдвигаемое возражение состоит в том, что с семантической точки зрения неправильно говорить *как то*, что «Джон знает, что ему больно», *так и то*, что «Джон не знает, что ему больно», хотя первое предложение определенно менее ошибочно, чем второе, поскольку оно удовлетворяет по крайней мере одному из условий, предполагаемых знанием, — Джон в состоянии утверждать, что ему больно. (На самом деле, здесь присутствуют и элемент *истинности* и элемент *уверенности*; трудности связаны с элементом подтверждения.)

Я не намерен рассматривать здесь эту проблему¹², поскольку меня больше заботит сходство между нашими двумя вопросами.

¹¹ Например, я знаю, что Солнце находится на расстоянии в 93 миллиона миль от Земли, но у меня нет *подтверждения* того, что это так. На самом деле, я даже не помню, откуда я узнал об этом.

¹² По сути, невозможно решить, девиантно или нет предложение «Джон знает, что ему больно», пока не сформулировано иначе условие подтверждения; сформулировано так, чтобы можно было избежать возражения, приведенного в сноске 11 (если вообще его можно переформули-

Например, мы могли бы принять (как «недевиантные», «логически нормальные», «несамопротиворечивые» и т. п.) следующие два утверждения:

- (a) Машина установила, что она находится в состоянии A ,
- (b) Джон знал, что ему больно,

или же отвергнуть их. Пусть мы отвергаем утверждения (a) и (b), но в этом случае мы можем найти альтернативные им формулировки, которые будут приемлемыми в семантическом отношении: например (для утверждения (a)), «Машина находилась в состоянии A , по этой причине она напечатала: «Я нахожусь в состоянии A »»; (для утверждения (b)), «Джон ощущал боль, по этой причине он сказал: «Мне больно»» (или «Джону было больно, и он показал это, сказав «Мне больно»»).

С другой стороны, если принять утверждения (a) и (b), то возникают вопросы: (a₁) «Как машина установила, что она находится в состоянии A ?» и (b₁) «Как Джон узнал, что ему больно?».

Если на эти вопросы вообще существуют ответы, то они будут вырожденными, например — «Находясь в состоянии A » и «Испытывая боль».

Теперь, я полагаю, совершенно ясно, что в обоих случаях трудности порождены одной и той же причиной, а именно: они вызваны тем обстоятельством, что «сообщение в словесной форме» («Я нахожусь в состоянии A » или «Мне больно») исходит непосредственно из состояния, о котором «сообщается»: чтобы получить «ответ», не требуется никакого «вычисления», никакого дополнительного «подтверждения». Все разногласия философов о том, «как следует говорить», по сути, связаны с поиском терминологии для описания когнитивных процессов вообще, что имеет место и в данном конкретном случае. (Заметьте, что традиционный в эпистемологии ответ на вопрос (b₁) — «путем интроспекции» — не согласуется в данном случае с фактами, поскольку он определенно предполагает наличие некоторого ментального события («акта» интроспекции), отличного от ощущения боли.)

В заключение предположим, что машина снабжена устройством для «сканирования» соседней машины T_1 . Тогда вопрос: «Как T

рывать так, чтобы хотя бы отчасти сохранилось это условие). Тем не менее, думаю, что дискуссия, приведенная выше, показывает, почему *желательно* считать это предложение девиантным.

устанавливает, что T_1 находится в состоянии A ?» может быть совершенно разумным, столь же разумным, как и вопрос: «Как T устанавливает, что трехтысячный разряд π — такой-то?» В обоих случаях ответ будет включать описание всей «программы» (а при необходимости и разъяснение рациональных соображений, лежащих в основе программы). Более того, потребуется не просто описать внутренние состояния T , но и охарактеризовать еще, как физически реализовали связи между T и T_1 (устройство органов чувств и т. д.), поскольку в данном случае T отвечает не на математический, а на эмпирический вопрос. Точно также вопрос: «Как Шерлок Холмс узнал, что Джону больно?» может быть совершенно разумным, а ответ на него может быть довольно сложным.

3. «МЕНТАЛЬНЫЕ» И «ЛОГИЧЕСКИЕ» СОСТОЯНИЯ

Рассмотрим два вопроса:

- (1) Откуда Джон знает, что ему больно?
- (2) Откуда Джон знает, что у него жар?

Как мы видели в предыдущем разделе, первый вопрос несколько странен, а второй вопрос может быть вполне разумным. Действительно, когда Джон говорит «Мне больно», ему никто не возразит: «Вы ошибаетесь». (Ему могли бы возразить: «Вы оговорились» или «Вы лжете», но не «Вы ошибаетесь».) С другой стороны, когда Джон говорит: «У меня жар», врач, только что измеривший ему температуру, вполне может возразить: «Вы ошибаетесь». И он может не иметь в виду, что Джон совершил лингвистическую ошибку, солгал или что-то напутал.

Можно было бы заключить, что для различия между утверждениями о своем собственном состоянии и утверждениями о состоянии других имеется аналог в случае машин, тогда как для затронутого сейчас различия между утверждениями о «ментальном» состоянии человека и утверждениями о его «физическом» состоянии, если воспользоваться традиционным языком, никакого аналога нет. Но это не так. Каков этот аналог, мы и попытаемся сейчас сформулировать.

Прежде всего, вернемся к понятию машины Тьюринга. При описании машины Тьюринга с помощью «машинной таблицы» указывается, что она имеет ленту, печатающее устройство и «сканирующее» устройство (оно может представлять собой простой заост-

ренный стержень машины, который в любой данный момент времени нацелен только на одну клетку ленты) и конечное множество «состояний» (*A, B, C* и т. д.). (В последующем, мы будем иногда называть их *логическими состояниями*, чтобы отличать от некоторых других состояний, которые мы введем чуть позже.) Помимо этого, при описании машины Тьюринга задаются только правила, жестко определяющие, в каком порядке состояния следуют друг за другом и что когда печатается.

В частности, «логическое описание» машины Тьюринга не содержит никаких данных о *физической природе* этих «состояний» — и даже о физической природе машины в целом. (Неизвестно, состоит ли она из электронных реле, картона, клерков, сидящих за пультами, или чего-либо еще.) Иными словами, любая данная «машина Тьюринга» — это *абстрактная* машина, которая может иметь практически неограниченное число различных физических реализаций.

Однако, как только машина Тьюринга получает физическую реализацию, происходит нечто интересное. Хотя с точки зрения логика машина обладает только состояниями *A, B, C* и т. д., для инженера она имеет практически неограниченное число дополнительных «состояний» (разумеется, в ином смысле этого слова — эти состояния мы будем называть *структурными состояниями*). К примеру, если машина состоит из вакуумных трубок, то может так случиться, что одна из ее трубок выйдет из строя, и благодаря этому машина перейдет в иное «состояние», — иное с точки зрения физика, но отнюдь не логика. С другой стороны, если машина сделана из картона и управляется вручную, и если ее картон может покоробиться, то это, очевидно, составляет одно из ее возможных «нелогических» или «структурных» состояний. И так далее.

Физически реализованная машина Тьюринга может не иметь возможности определить свое собственное структурное состояние, точно так же, как человек может не знать о состоянии своего аппендикса в данный момент времени. Однако, очень удобно снабдить машину электронными «органами чувств», которые позволят ей сканировать саму себя и обнаруживать малейшие неисправности. Мысленно можно представить себе, что эти «органы чувств» вызывают печатание определенных символов на «входной ленте», которую машина время от времени «обследует». (Небольшое затруднение возникает из-за того, что «сообщение» от органов чувств может занимать несколько клеток ленты, тогда как машина «сканирует» только одну клетку за раз. Однако, это несуществен-

но, поскольку хорошо известно, что можно достичь эффекта «чтения» любого конечного числа клеток с помощью программы, которая предполагает сканирование только одной клетки за раз.)

(В качестве отступления отмечу, что первые сконструированные цифровые компьютеры не имели устройств описанного вида. С другой стороны, они *содержали* свыше 3000 вакуумных трубок, и некоторые из этих трубок в какой-то момент выходили из строя! Поэтому очень быстро была осознана необходимость «программ самоконтроля»)¹³

Машина, способная распознавать, по крайней мере, часть собственных структурных состояний, находится приблизительно в таком же положении, что и человек, который может определять с различной степенью достоверности только некоторые дисфункции своего организма. Так, предположим, что машина «печатает»: «Вышла из строя вакуумная трубка 312». Вопрос «Как машина установила, что вышла из строя вакуумная трубка 312?» совершенно разумен. А ответ может содержать ссылку как на физическую структуру машины («органы чувств» и т. п.), так и на ее «логическую структуру» (программу «чтения» и «интерпретации» входной ленты).

Если машина печатает: «Вышла из строя вакуумная трубка 312», хотя та на самом деле работает, ошибка могла произойти или в результате неправильного вычисления (в процессе «чтения» и «интерпретации» входной ленты) или вследствие неправильного сигнала, поступившего от органов чувств. С другой стороны, если машина печатает: «Я нахожусь в состоянии А» и делает это просто потому, что ее машинная таблица содержит команду: *Напечатать: «Я нахожусь в состоянии А»*, то вопроса о неправильном вычислении возникнуть не может. Даже если из-за какого-то сбоя печатающий механизм печатает: «Я нахожусь в состоянии А», хотя машина *не* находится при этом в состоянии А, это не будет «неправильным вычислением» (а только, так сказать, «обмолвкой»).

Интересно отметить, что как существуют два возможных описания поведения машины Тьюринга — структурный чертеж инженера и «машинная таблица» логика, так существуют и два возможных описания человеческой психики. «Бихевиористский» подход (к

¹³ В действительности не потребовалось добавлять никаких «органов чувств»; существующие компьютеры проверяют себя, «проводя над собой решающие эксперименты» (т. е. выполняя определенные проверочные вычисления и сравнивая полученные результаты с заданными правильными результатами).

этой категории относятся и теории, использующие «гипотетические построения», в частности «построения» на основе физиологии) имеет своей целью, в конечном счете, полное физикалистское¹⁴ описание всего человеческого поведения в терминах химии и физики. Это соответствует описанию физически реализованной машины Тьюринга, сделанному инженером или физиком. Но можно было бы попытаться дать и более абстрактное описание человеческих ментальных процессов в терминах «ментальных состояний» (когда физическая реализация не уточняется) и «восприятий» (impressions) (последние выполняют роль символов на машинных лентах). Это описание определяло бы законы, устанавливающие порядок следования состояний друг за другом и их отношение к вербализации (или, во всяком случае, к вербальному мышлению). Это описание, которое было бы аналогом «машинной таблицы», и есть, по существу, программа, выдвинутая классической психологией! Зачастую считают, что классическая психология потерпела неудачу по методологическим причинам. В свете предложенной аналогии я бы счел, что, скорее, она потерпела неудачу по эмпирическим причинам — ментальные состояния и «восприятия» людей не образуют такую же причинно замкнутую систему, какая имеет место в случае «конфигураций» машины Тьюринга.

Проведенная аналогия между логическими состояниями машины Тьюринга и ментальными состояниями человека, с одной стороны, и структурными состояниями машины Тьюринга и физическими состояниями человека, с другой, наводит, на мой взгляд, на серьезные размышления. В частности, дальнейшее исследование этой аналогии позволит глубже прояснить обсуждаемое нами понятие «ментального состояния». Это «исследование» еще не предпринималось, во всяком случае мною, но для тех, кого это может заинтересовать, я хотел бы указать несколько характеристик, которые, видимо, отличают логические и ментальные состояния от структурных и физических состояний, соответственно:

(1) Функциональную организацию (решение задач, мышление) человека или машины можно описывать, используя последовательность ментальных или логических состояний, соответственно (и

¹⁴ В смысле Оппенгейма (*Oppenheim P., Putnam H. Unity of Science as a Working Hypothesis // Feigl H., Maxwell G., Scriven M. (eds.) Minnesota Studies in the Philosophy of Science, 1958, Vol. II, Minneapolis, p. 3—36*); а не в том «эпистемологическом» смысле, который ассоциируется с тем, что писал о «физикализме» Карнап.

сопутствующие им вербализации), не обращаясь к природе «физической реализации» этих состояний.

(2) Состояния, видимо, внутренне связаны с *вербализацией*.

(3) Если мы имеем дело с рациональным мышлением (или вычислением), то «программа», определяющая последовательность состояний и т. д., открыта для рациональной критики.

4. «ТОЖДЕСТВО» СОЗНАНИЯ И МОЗГА

В заключение мы хотели бы сравнить людей и машины с точки зрения решения вопроса о *тождественности* ментальных состояний соответствующим физическим состояниям (или о тождественности логических состояний соответствующим структурным состояниям). Как отмечалось в начале статьи, все аргументы «за» и «против» подобного отождествления можно прекрасно сформулировать с помощью машин Тьюринга.

Витгенштейн, например, в 1930-х годах использовал следующий аргумент: если я имею в сознании некоторый остаточный образ и одновременно наблюдаю состояние своего мозга (с помощью соответствующего прибора), то я наблюдаю *две* вещи, а не одну. (По-видимому, это аргумент *против* отождествления.) Но мы вполне можем представить себе «умную» машину Тьюринга, «рассуждающую» следующим образом: «Когда я печатаю “Я нахожусь в состоянии А”, мне не нужно использовать свои “органы чувств”. Если же я действительно использую свои “органы чувств” и сравниваю случаи, когда я нахожусь в состоянии А, и случаи, когда взведен триггер Зб, то я сравниваю *две* вещи, а не одну». Не думаю, что нам покажется убедительным этот сформулированный на «машинный» манер аргумент Витгенштейна!

Рассел, напротив, довел до абсурда утверждения о «тождестве», заявив однажды, что все, что мы когда-либо *видели*, — это части нашего собственного мозга. Аналогичным образом, «Рассел-машина» мог бы «утверждать», что «все когда-либо мной наблюдаемое — это мои собственные вакуумные трубки». И тот, и другой «Рассел» неправы: человек наблюдает события во внешнем мире, но процесс «наблюдения» предполагает события в его мозгу. Сходным образом, машина *T* может «наблюдать», скажем, банки с томатным супом (если в ее задачу входит сортировка банок с супом), но процесс «наблюдения» предполагает функционирование вакуумных трубок. Но у нас нет никаких причин говорить, что «на самом деле» машина наблюдает свои собственные вакуумные трубки.

Рассмотрим, однако, более серьезные аргументы. В начале статьи я отмечал, что *синтетический* характер утверждения (1) «Я испытываю боль, если и только если возбуждены мои С-волокна» использовался в качестве аргумента в защиту той точки зрения, что такие «свойства» (или «события», или «состояния»), как «иметь возбужденными С-волокна» и «испытывать боль», не могут быть тождественными. Этот аргумент неудовлетворителен по крайней мере по двум причинам: (а) различие между «аналитическим и синтетическим» не является достаточно четким, особенно если речь идет о научных законах; и (б) весьма спорен используемый здесь критерий отождествления «свойств» («событий», или «состояний»).

Что касается пункта (а), то, как я показал в другой своей статье¹⁶, фундаментальные научные законы нельзя отнести ни к «аналитическим», ни к «синтетическим» высказываниям. Рассмотрим, к примеру, какой концептуальный сдвиг был вызван переходом от евклидовой геометрии к неевклидовой, или был бы вызван, если бы мы отказались от закона сохранения энергии. Было бы неверно сказать, что законы евклидовой геометрии (пока они господствовали) были «аналитическими», а Эйнштейн просто «изменил значение слов». В действительности же, именно потому, что Эйнштейн, *не* меняя значение слов, продолжал говорить о кратчайших траекториях в пространстве, в котором мы существуем, движемся и живем, общая теория относительности и показалась столь непостижимой, когда была впервые сформулирована. Утверждать, что можно вернуться в то же самое место, двигаясь по прямой в одном и том же направлении! Приняв общую теорию относительности, мы, по сути, приняли новую систему понятий, а не просто «новую систему слов-ярлыков».

Но если неверно уподоблять пересмотр фундаментальных научных законов принятию новых лингвистических конвенций, то в равной мере ошибочно и уподоблять — вслед за традиционными философами науки — концептуальное изменение, озаглавленное Эйнштейном, такому типу изменения, как изменение, вызванное обнаружением черного лебедя (прежде мы считали, что все лебеди белые)! Фундаментальные законы подобны принципам чистой математики (как подчеркивал Куайн), ибо их нельзя опровергнуть с помощью отдельных экспериментов: мы всегда можем сохранить

¹⁶ Putnam H. The Analytic and the Synthetic // Mind, Language and Reality, Philosophical Papers. Vol. 2. Cambridge: Cambridge University Press, 1975, p. 33—69.

законы и объяснить эксперименты, прибегнув к разнообразным *ad hoc* способам. Несмотря на уничижительный оттенок слова «*ad hoc*», поступать так в случае важных научных теорий даже *рационально, если нет приемлемой альтернативной теории*. Вот почему потребовалось растянувшееся на сто лет формирование новой системы понятий — а не только несколько экспериментов — чтобы ниспровергнуть евклидову геометрию. Сходным образом, именно поэтому мы не можем сегодня описать *никаких* экспериментов, которые *сами по себе* опровергали бы закон сохранения энергии, хотя этот закон не является «аналитическим» и мог бы быть отброшен, если бы новому Эйнштейну удалось выдвинуть серьезные *теоретические* основания для этого, подкрепленные подтверждающими экспериментами.

Как сказал Хэнсон¹⁶, наши понятия имеют «встроенные в» них теории, поэтому отказ от основной научной теории без выдвижения альтернативной вызвал бы «разрушение нашей системы понятий». С другой стороны, хотя мы и *могли бы* вопреки фактам продолжать *считать*, что «все лебеди белые», в этом не было бы никакого *смысла*, ибо в этом случае понятия не *связаны* с принятием одного из противоположных принципов — так, как связаны геометрические понятия с принятием *какой-либо* геометрии, пусть и необязательно евклидовой.

Не отрицаю, что *сегодня* любая обнаруженная «корреляция», имеющая форму: «Некто находится в ментальном состоянии ψ , если и только если его мозг находится в состоянии ϕ », *вначале* была бы *просто* корреляцией, чисто «эмпирическим обобщением». Но, на мой взгляд, интерес представляет случай, который имел бы место, если бы нам удалось сформулировать и теоретически разработать *систему* подобных «корреляций». В этом случае значительно изменился бы научный язык. Ученые стали бы говорить: «Невозможно *в принципе* находиться в ментальном состоянии ψ , чтобы мозг не находился в состоянии ϕ ». И очень может быть, что «принципиальная невозможность» означала бы то, что Хэнсон справедливо называет *концептуальной* невозможностью (ср. Hanson, 1958): ученые не могли бы *вообразить* (за исключением нового Эйнштейна), что некто находится в ментальном состоянии ψ , а его мозг не находится в состоянии ϕ . В частности, ни один эксперимент не мог бы *сам по себе* опровергнуть психологические законы,

¹⁶ Hanson N. Patterns of Discovery. Cambridge: Cambridge University Press, 1958.

получившие такого рода статус¹⁷. Не ясно ли, что в подобной научной ситуации было бы некорректно утверждать, что ϕ и ψ — это одно и то же состояние?

Более того, вовсе неясны критерии для отождествления «событий», «состояний» или «свойств». В качестве примера закона, имеющего обсуждаемый нами статус, приведем следующий: Свет проникает через апертуру, если и только если через эту апертуру проникает электромагнитное излучение (имеющее такую-то длину волны).

Очевидно, что этот закон *не* является «аналитическим» утверждением. Тем не менее с научной точки зрения было бы вполне корректно говорить, что: (i) свет, проникающий через апертуру, и (ii) электромагнитное излучение (имеющее такую-то длину волны), проникающее через апертуру, — это два описания одного и того же события. (Фактически, в «обыденном языке» не только не требуется, чтобы описания одного и того же явления были эквивалентными; но даже можно говорить о *несовместимости* описаний одного и того же события!)

Впрочем, можно было бы возразить, что *свойства* (в противоположность событиям) нельзя описывать с помощью различных неэквивалентных описаний. По сути, Фреге, Льюис и Карнап *отождествили* свойства и «значения» (в том смысле, что, если *по определению*, два выражения имеют разные значения, то они «обозначают» разные свойства). На мой взгляд, это весьма сомнительно. Но предположим, что это верно. Что из этого следовало бы? Например, пришлось бы признать, что испытать боль и иметь возбужденными С-волокна — это разные свойства. Однако, если взять язык машины Тьюринга, описанной в начале статьи и способной «строить теории», то пришлось бы в равной мере признать, что «находиться в состоянии А» и «иметь взведенным триггер 3б» — это разные свойства. Действительно, на языке машины предложения (i) «Я нахожусь в состоянии А» и (ii) «Взведен триггер 3б» согласно любому критерию не являются синонимами (они имеют разные синтаксические свойства и разные «условия произнесения» (conditions of utterance), например, машине приходится использовать разные «методы верификации»). Поэтому тот, кто пожелает на

¹⁷ См. обсуждение геометрии в статье: Putnam H. The Analytic and the Synthetic // Mind, Language and Reality, Philosophical Papers. Vol. 2. Cambridge: Cambridge University Press, 1975, p. 33—69.

этом основании доказывать существование души, должен быть готов принять в свое философское лоно и души машин Тьюринга.

5. «ЛИНГВИСТИЧЕСКИЙ» АРГУМЕНТ

Последним в вопросе о тождестве сознания и тела я рассмотрю широко используемый «лингвистический» аргумент; его, в частности, выдвинул Макс Блэк против Герберта Фейгля на конференции, результатом которой стала книга «Измерения мышления»¹⁸. Рассмотрим предложение:

(1) Боль тождественна возбуждению С-волокон.

Это предложение девиантно (так строится этот аргумент, хотя терминология используется другая), ибо не существует такого утверждения, для выражения которого оно могло бы использоваться в обычных условиях. Поэтому, если философ выдвигает его в качестве тезиса, он, стало быть, придает этим словам новое значение, а не высказывается о каком бы то ни было открытии. Например (утверждал Макс Блэк), можно было бы начать говорить: «У меня возбуждены С-волокна» вместо — «Я испытываю боль» и т. п. Но в этом случае мы бы *просто* присвоили выражению «иметь возбужденными С-волокна» новое значение «испытывать боль». Это означает, что пока слова сохраняют свои нынешние значения, предложение (1) остается неосмысленным (unintelligible).

Я согласен, что предложение (1) является «девиантным» в современном языке. Но я *не* согласен с тем, что предложение (1) не сможет стать нормальным, недевиантным предложением, пока не изменятся нынешние значения слов.

Если выразить суть в двух словах, то, что считать «девиантным», во многом определяется контекстом, который включает и состояние нашего знания; с развитием новых научных теорий постоянно происходит то, что предложения, которые прежде «не употреблялись», которые прежде были «девиантными», входят в употребление, но не потому, что слова приобретают *новые* значения, а потому что старые значения, образующие ядро обычных употреб-

¹⁸ Dimensions of Mind // Proceedings of the IIIrd Annual New-York University Institute of Philosophy. May 15—16, 1959. New York University Press, 1960.

лений этих слов, *обуславливают* (determine) новый способ их употребления с учетом нового контекста.

Нет ничего предосудительного в том, чтобы попытаться использовать лингвистическую теорию для решения этой проблемы, но она должна быть довольно сложной. Реально вопрос стоит не для *синхронической*, а для *диахронической*¹⁹ лингвистики, т. е. речь идет не о том, «девиантно ли *сегодня* предложение (1)?», а о том, «обязательно ли изменились бы значения слов, если бы изменения в научном знании (например, появление интегрированной сети “высокоприоритетных” психологических законов в нашем общем научном мировоззрении) привели к тому, что предложение (1) стало бы *недевиантным* предложением», а это не такой уж простой вопрос.

Хотя сейчас не время и не место заниматься разработкой семантической теории²⁰, рискну сделать несколько замечаний по этому вопросу.

Во-первых, нетрудно показать, что простое произнесение предложения, которое прежде никем не произносилось, необязательно вводит «новый способ употребления». Если я говорю: «На этом столе лежит ящерица-ядозуб», то, видимо, я произношу предложение, которое до меня никто не произносил, но я никоим образом не изменяю значение используемых слов.

Во-вторых, даже если предложение, прежде девиантное, начинает обретать стандартное употребление, то это вовсе не означает, что произошло изменение *значения* используемых слов. Так, предложение «Я нахожусь за тысячи миль от вас» или его древнегреческий перевод, несомненно, были девиантными до изобретения письменности, но обрели («*обрели*», а не им «*задали*») нормальное

¹⁹ Диахроническая лингвистика изучает язык с учетом его изменений во времени; синхроническая лингвистика направлена на описание языка в данный конкретный момент времени.

²⁰ Более подробное изложение см. в работе Зиффа (*Ziff P. Semantic Analysis*, N. Y., 1960). Я глубоко признателен Зиффу и за то, что эта работа попала в мои руки, и за его личное обсуждение со мной этих проблем. Раздел 5 настоящей статьи отчасти отражает влияние Зиффа (особенно это касается различия между «синхроническим и диахроническим»), а отчасти служит приложением некоторым идеям, выдвинутым мной в эту тему в статье: *Putnam H. The Analytic and the Synthetic // Mind, Language and Reality, Philosophical Papers. Vol. 2. Cambridge: Cambridge University Press, 1975, p. 33—69.*

употребление с изобретением письменности и появлением возможности личной переписки на больших расстояниях.

Отметим причины, почему мы не считаем, что какое-либо слово (например, «я», «ты», «тысяча») в этом предложении изменило свое значение: (а) новое употребление не было *случайным*, не было *результатом соглашения*; оно представляло собой автоматическую проекцию ²¹ имевшихся обычных употреблений тех нескольких слов, что составляют это предложение, на новый контекст; (б) значение предложения в целом является функцией значений отдельных слов, из которых оно составлено. (Фактически, этот принцип лежит в основе всего понятия значения слова — так, если мы говорим, что предложение изменило свое значение, то мы должны ответить на вопрос: «*Какое слово изменило свое значение?*» Совершенно очевидно, что в рассматриваемом случае этот вопрос поставил бы нас в тупик.)

В рассмотренном примере новый контекст возник в результате появления новой технологии, но сходное влияние на язык может оказывать и новое теоретическое знание. (Например, предложение «он обошел вокруг Земли» девиантно в той культуре, где неизвестно, что Земля круглая!) Сходный случай описывает Малкольм: благодаря различным физиологическим показателям (электроэнцефалограммам, движению глаза во время сна, перепадам кровяного давления и т. п.) у нас появляется возможность определять, когда начинаются и заканчиваются сны. Таким образом, предложение «Он достиг середины своего сна» когда-нибудь обретет стандартное употребление. Малкольм прокомментировал это таким образом, что словам в этом случае будто бы будет задано некоторое новое употребление. Думаю, Малкольм ошибается; в этом случае предложение обретет употребление *благодаря* значению входящих в него слов — в противоположность случаю, когда словам в буквальном смысле *задают* некоторое употребление (например, устанавливают соглашения о значении выражений). Думаю, совершенно правильно «реалистическое» объяснение этого случая: данное предложение прежде не употреблялось, потому что у нас не было способа определить, когда начинаются и прекращаются сны. Теперь у нас появляются такие способы, и поэтому мы находим ситуации, когда естественно употребить это предложение. (Отметим, что у Мал-

²¹ Термин позаимствован у Зиффа (*Ziff P. Semantic Analysis, N. Y., 1960*).

кольма нет объяснения тому факту, что мы задаем *этому* предложение именно *это* употребление.)

Впрочем, кто-то может признать, что изменения значения не следует путать с изменениями в распределении (distribution) слова ²², и что научные и технологические успехи часто порождают последние изменения, которые некорректно считать изменениями значения. Но при этом он мог бы утверждать, что хотя можно бы было заранее предугадать условия, при которых предложение «Он обошел вокруг Земли» станет недевиантным, мы сегодня не способны предугадать условия, при которых ²³ станет недевиантным предложение «Ментальное состояние ψ тождественно состоянию мозга ϕ ». Но это не очень серьезное возражение. Во-первых, вполне могло бы быть, что первобытные люди были абсолютно неспособны представить шарообразную Землю (ведь тогда люди на «обратной» стороне Земли должны были бы падать с нее). Даже сорок лет назад было трудно, а то и невозможно, представить условия, при которых предложение «Он достиг середины сна» было бы недевиантным. И, во-вторых, я полагаю, что *можно* в общем виде описать условия, при которых предложение «Ментальное состояние ψ тождественно состоянию мозга ϕ » стало бы недевиантным.

Для выполнения этой задачи нам нужно рассмотреть одну важную разновидность связки «есть», используемую при *теоретическом отождествлении*. Указанная связка «есть» используется в следующих предложениях:

(2) Свет есть электромагнитное излучение (с такими-то длинами волн).

(3) Вода есть H_2O .

Когда, например, научное сообщество приняло отождествление (2), это означало приблизительно следующее: до отождествления имелись две отдельные теории — оптическая теория (существо которой великолепно раскрыл Тулмин в своей книге по философии науки) и электромагнитная теория (представленная уравнениями Максвелла). Решение *определить* свет как «электромагнитное излучение с такими-то длинами волн» в научном плане (как уже не раз отмечалось) оправдывалось следующими соображениями:

²² *Распределение* слова — это множество предложений, в которых оно встречается.

²³ Предложение «Ментальное состояние ψ тождественно состоянию мозга ϕ » используется здесь как заменитель таких предложений как «Большое тождественно возбуждению С-волокон».

(1) Это сделало возможным *выведение* законов оптики (в первом приближении) из более «фундаментальных» законов физики. Таким образом, даже если бы это теоретическое отождествление не дало никаких других результатов, оно способствовало бы упрощению структуры научных законов.

(2) Это сделало возможным получение *новых* предсказаний в «редуцируемой» теории (т. е. в оптике). В частности, ученые смогли предсказать, что в определенных случаях *не* будут выполняться законы геометрической оптики. (Ср. известные замечания Дюгема о сведении законов Кеплера к законам Ньютона.)

Попробуем теперь описать условия, при которых теоретическое отождествление ментальных состояний с физиологическими могло бы быть выполнено как надежная научная процедура. Говоря обобщенно, нам нужны не *просто* «корреляты» для субъективных состояний, а нечто более тщательно изученное, например, мы должны знать о физических состояниях то, что позволило бы нам не *просто предсказывать* человеческое поведение, но и давать ему каузальное объяснение (скажем, мы должны знать микросостояния основных процессов).

Во избежание «категориальных ошибок» (category mistakes), нужно очень четко определить, что значит «объяснить человеческое поведение». Предположим, человек говорит: «Я плохо себя чувствую». При одной системе категорий, его поведение можно описать так: «Он утверждает, что плохо себя чувствует». А объяснение возможно такое: «Он сказал, что плохо себя чувствует, потому что он голоден и у него болит голова». Я вовсе не хочу сказать, что событие «Джон *утверждает*, что плохо себя чувствует» можно объяснить с помощью законов *физики*. Но к этому событию имеет отношение *еще одно* событие, а именно — «Тело Джона генерирует такие-то звуковые волны». С одной точки зрения, это — «другое событие», по сравнению с событием «Джон утверждает, что плохо себя чувствует». Однако (перелагая замечание Хэнсона) было бы бессмысленно говорить, что это разные события, если бы в каком-то смысле они не были *тем же самым* событием. А для нас здесь как раз и важно то, в каком смысле они являются «тем же самым событием», а не то, в каком смысле они — «разные события».

Словом, когда я говорю о «каузальном объяснении человеческого поведения», все, что я имею в виду, — это каузальное объяснение определенных физических событий (движений тела, генерации звуковых волн и т. п.), которые «тождественны» в только

что описанном смысле событиям, составляющим человеческое поведение. И никакие аргументы в духе Райла не смогут исключить²⁴ ту очевидную возможность, что физика могла бы добиться многого в подобном объяснении.

Если бы это уже произошло, то теоретическое отождествление «ментальных состояний» с их «коррелятами» имело бы следующие два преимущества:

(1) Это сделало бы возможным (опять-таки в «первом приближении») выведение из законов физики классических законов (или обобщений низшего уровня) «менталистской» психологии здравого смысла, таких как: «Люди стремятся избегать то, с чем связаны болезненные переживания».

(2) Это позволило бы предсказывать случаи (а их множество), которые «менталистская» психология здравого смысла не способна предсказать.

Конечно, преимущество (2) можно было бы получить и без «отождествления» (используя законы установления корреляций). Но преимущество (2) в равной мере можно было бы получить без отождествления и в случае оптики (предположив, что свет *сопровождается* электромагнитным излучением, но не тождествен ему.) Однако ни в том, ни в другом случае никаким иным способом, кроме отождествления, нельзя было бы *одновременно* элиминировать несколько законов (в пользу теоретических определений) и при этом увеличить объяснительную силу теории. Следует заметить, что любой аргумент «за» и «против» отождествления равным образом применим как в случае, когда речь идет о сознании и мозге, так и в случае, когда речь идет о свете и электромагнитном излучении. (Против отождествления света и электромагнитного излучения можно было бы выдвинуть даже аргумент на основе «обыденного языка».)

Два небольших замечания: (i) Когда я называю предложение «свет есть электромагнитное излучение (с такими-то длинами волн)» определением, я не имею в виду, что оно является «аналитическим». Но в таком случае «определения» *в собственном смысле слова* фактически никогда в теоретической науке не быва-

²⁴ Что попытался сделать один молодой философ в недавней статье в журнале «British Journal for the Philosophy of Science».

ют аналитическими ²⁶. (Куайн заметил однажды, что ему известны по крайней мере девять приемлемых значений «определения» и ни в одном из этих значений определение не является аналитическим.) Конечно, философ мог бы тогда поставить под сомнение *рациональную оправданность* теоретического отождествления, сославшись на, что мы ничего не достигаем, элиминировав «законы» в пользу «определений», раз те и другие — *синтетические* предложения. Тот факт, что ученый вовсе так не думает, лишний раз показывает, сколь бесполезно смотреть на науку под углом зрения «аналитического и синтетического». (ii) Принятие теоретического отождествления, такого, как, например, «Боль есть возбуждение С-волокон», не обязывает нас, как считал Блэк, *употреблять* термины «боль» и «возбуждение С-волокон» как взаимозаменяемые в разговорных выражениях. Например, отождествление «воды» с « H_2O » сегодня хорошо известно, но никто не говорит (разве только в шутку): «Принесите мне стакан H_2O ».

Думаю, что предложенные здесь соображения позволяют: (a) объяснить тот факт, что в современном языке предложения типа «Ментальное состояние ψ тождественно состоянию мозга ϕ » девиантны, и в то же время (б) показать, каким образом эти же предложения могли бы стать недевиантными при соответствующем углублении наших научных знаний о физической природе и причинах человеческого поведения. Рассматриваемые предложения нельзя сегодня использовать как предложения, выражающие теоретическое отождествление, ибо такое отождествление не было выполнено. Теоретическое отождествление — это не то действие, которое можно выполнить «по желанию»: его выполнение, как и выполнение многих других действий имеет *предпосылки*, которые сегодня еще не сложились. С другой стороны, если бы появилась описанная выше научная теория, то возникли бы и предпосылки для теоретического отождествления, как это имело место в случае света и электромагнитного излучения, и тогда предложения рассматриваемого вида *автоматически* вошли бы в употребление — как выражения для соответствующих теоретических отождествлений. Повторяю: этот способ *обретения* употребления отличается от случая, когда употребление задается (и от «изменения значения» в собственном смысле слова), тем, что «новое употребление» пред-

²⁶ Этот вопрос обсуждается мной в статье: Putnam H. The Analytic and the Synthetic // Mind, Language and Reality, Philosophical Papers. Vol. 2. Cambridge: Cambridge University Press, 1975, p. 33—69.

представляет собой автоматическую проекцию существующих употреблений и не предполагает произвольно принимаемого соглашения (если не считать элемент «соглашения», связанный с принятием *любой* научной гипотезы, в том числе и гипотезы «Земля круглая»).

До сих пор мы рассматривали только предложения, имеющие форму ²⁶ «Ментальное состояние ψ тождественно состоянию мозга ϕ ». А как быть с предложением:

(3) Ментальные состояния есть микросостояния мозга?

Это предложение не «задает», так сказать, никакого *конкретного* теоретического отождествления: оно говорит только о возможности теоретического отождествления, которое не конкретизируется. Именно такого рода утверждение мог бы высказать Фейгл. Блэк ²⁷ мог бы в ответ указать, что, произнося (3), Фейгл произнес бы странный набор слов (т. е. девиантное предложение). Возможно, Блэк прав. Возможно, что предложение (3) девиантно в современном языке. Но возможно и то, что наши потомки через два или три столетия будут считать, что Фейгл был совершенно прав и то лингвистические возражения, выдвинутые против (3), были дольно нелепыми. И может оказаться, что они будут правы.

6. МАШИННАЯ ЛИНГВИСТИКА

Рассмотрим сформулированный выше лингвистический аргумент с точки зрения предлагаемой в настоящей статье аналогии между человеком и машиной Тьюринга. Как мы увидим, наша машина Тьюринга, вероятно, не сможет, при отсутствии соответствующих «органов чувств», построить правильную теорию о своем собственном строении. С другой стороны, предложение «Я нахожусь в состоянии *A*» будет с определенной регулярностью появляться в машинном «языке». Если машинный «язык» будет достаточно сложен, то с помощью синтаксического анализа можно будет выявить конечное множество основных составных блоков (морфем) и описать правила построения из них потенциально бесконечного множества «предложений». В частности, мы сможем различать

²⁶ Под предложениями этой *формы* я имею в виду предложения, образованные в результате разных *подстановок* в предложение «Ментальное состояние ψ тождественно состоянию мозга ϕ ». См. сноску 23.

²⁷ С некоторыми колебаниями я приписал эту позицию Блэку на основании его замечаний на одной из конференций. Но я, конечно, понимаю, что он не несет всей ответственности за замечания, сделанные мимоходом.

*грамматически правильные*²⁸ и *грамматические неправильные предложения* в машинном «языке». Сходным образом, можно будет установить, с какой регулярностью появляется то или иное предложение (или, используя оксфордский жаргон, «описать способы употреблений предложений») и, приписывая «значения» конечному множеству морфем и конечному множеству правил построения, эффективно определить «способы употребления» разнообразных предложений как проекции значений отдельных морфем и правил построения. В этом случае в «машинном языке» можно было бы различать не только «грамматически правильные» и «грамматически неправильные» предложения, но и «девиантные» и «недевиантные» предложения.

Чизом настаивал бы на том, что неправомерно приписывать машинам использование языка, и я с ним согласен. По этой причине я изредка заключаю в кавычки слова «язык», «значение» и т. д. — чтобы подчеркнуть там, где это необходимо, что эти слова используются в расширенном смысле. С другой стороны, важно осознать, что лингвистическое поведение машины может быть полным *аналогом* языка, и, поэтому, к нему целиком применима лингвистическая теория. Если читатель хочет убедиться в этом, ему следует внимательно просмотреть работу Хомского «*Синтаксические структуры*»²⁹, где ни в одном месте не предполагается, что совокупность высказываний, изучаемых лингвистом, была произнесена организмом, обладающим сознанием. Затем читателю следует обратиться к новаторской работе Зиффа «*Семантический анализ*»³⁰, посвященной эмпирической семантике, и он убедится, что это верно и в отношении *семантической* теории.

В этой связи выскажу еще два замечания: (i) поскольку я утверждаю, что психофизическая проблема *аналогична в строгом смысле* проблеме отношения между структурными и логическими состояниями, но не *тождественна* ей, то нам достаточно и соответствующей *аналогии* (а не тождества) между машинным «языком» и человеческим языком. (ii) Чизом мог бы возразить, что «бихевиористская» семантика разрабатываемого Зиффом вида (т. е. семантика, где «интенциональность» не выступает исходным понятием) невозможна. Но даже если бы это было так, это не имело бы

²⁸ Этот термин используется не в традиционном смысле, а так, как его трактовал Хомский (1957 (см. сноску 29)).

²⁹ Chomsky N. Syntactic Structures, The Hague, 1957.

³⁰ Ziff P. Semantic Analysis, N. Y., 1960.

значения для нас. Если *какая-либо* семантическая теория годится как описание человеческого языка, то нужно еще доказать, почему полностью *аналогичная* ей теория не годится как описание языка соответствующей машины. Например, если «интенциональность» служит исходным понятием в *научном* объяснении человеческого языка, то в случае машинного «языка» *такую же* объяснительную роль будет выполнять теоретическая конструкция, имеющая сходные *формальные* отношения с соответствующими «наблюдаемыми сущностями».

Конечно, кто-то мог бы истолковать возражение против «бихевиористской» лингвистики как аргумент, направленный *по сути* против любой попытки создать *научную* лингвистику. Но я не считаю оправданным допускать такую возможность.

Предположим, что мы снабдили нашу «создающую теории» машину Тьюринга «органами чувств», и, поэтому, она может получать эмпирические данные, необходимые для построения теории о своей собственной природе.

Теперь она может вводить в свой «теоретический язык» имена, «переводимые», скажем, как «триггер 36», и предложения, переводимые как «взведен триггер 36». Эти выражения в машинном языке будут иметь другое значение и другое употребление по сравнению с выражением «Я нахожусь в состоянии А».

Если какой-либо «лингвистический» аргумент действительно доказывает, что предложение «Боль тождественна возбуждению С-волокон» девиантно в современном языке, то согласно такому же аргументу, в машинном языке должно быть девиантным предложение «Состояние А тождественно взведенному состоянию триггера 36». Если согласно какому-то аргументу предложение «Боль тождественна возбуждению С-волокон» не может стать недевиантным (теперь язык рассматривается в *диахроническом* срезе) до тех пор, пока слова не изменят своих значений, то такой же аргумент в «диахронической лингвистике машинного языка» доказывал бы, что предложение «Состояние А тождественно взведенному состоянию триггера 36» не может стать недевиантным в машинном языке до тех пор, пока слова не изменят своих значений. Короче говоря, все философские аргументы, когда-либо выдвигаемые в связи с психофизической проблемой — как наиболее ранние и наивные (как, например, аргумент, гласящий, что «различие между состояниями сознания и физическими состояниями можно *видеть*»), так и наиболее изощренные, — имеют свой точный аналог, когда мы рас-

смотрим «проблему» об отношении логических и структурных состояний машин Тьюринга.

7. ЗАКЛЮЧЕНИЕ

Думаю, мораль ясна: невозможно больше считать, что психофизическая проблема представляет собой настоящую теоретическую проблему и ее «решение» могло бы пролить хоть какой-нибудь свет на мир, в котором мы живем. Ибо очевидно, что ни один взрослый человек в здравом уме не стал бы серьезно относиться к проблеме «тождественности» или «нетождественности» логических и структурных состояний машины — не потому, что ответ здесь очевиден, а потому что совершенно неважно, каков *ответ*. Но если так называемая «психофизическая проблема» — это просто иной способ осмыслить то же самое множество логических и лингвистических вопросов, то она, стало быть, столь же пуста и в такой же мере касается слов.

Часто очень важно осознать, что две проблемы, относящиеся к разным предметам исследования, совпадают во всех своих логических и методологических аспектах. В рассматриваемом случае нам нужно осознать, что любой вывод, полученный в ходе анализа психофизической проблемы, можно было бы получить, опираясь *на те же самые основания*, и в отношении машин Тьюринга. Но возьмем, к примеру, вывод о том, что логические и структурные состояния машин Тьюринга безнадежно различны. Если нам ясно (а это ясно со всей очевидностью), что этот вывод, даже будучи правильным, был бы простым открытием *о словах* (verbal discovery), то применительно к человеку такой же вывод, *полученный на основе таких же аргументов*, так же должен быть открытием о словах. Иными словами, если считать, что психофизическая проблема представляет не просто концептуальный интерес (например, она связана с вопросом о том, есть ли у людей «душа»), то, стало быть, *или (а)* ни один философ *никогда* не выдвинул аргумента, проливающего свет на решение этой проблемы (неважно, какую позицию он защищал), *или (б)* правилен какой-то философский аргумент в пользу механицизма, *или (в)* какой-то дуалистский аргумент действительно доказывает, что *и* люди, *и* машины Тьюринга обладают душой! Я предоставляю читателю самому решить, какая из этих трех альтернатив похожа на правду.

ПСИХОЛОГИЧЕСКИЕ ПРЕДИКАТЫ (ПРИРОДА МЕНТАЛЬНЫХ СОСТОЯНИЙ) *

Как правило, философа, изучающего сознание, интересует три вопроса: (1) Откуда мы знаем, что другие люди испытывают боль? (2) Является ли боль состоянием мозга (*brain state*)? (3) Что представляет собой анализ понятия *боли*? В этой статье я не буду обсуждать вопросы (1) и (3). Я выскажусь по вопросу (2) ¹.

1. ВОПРОСЫ ТОЖДЕСТВА

«Является ли боль состоянием мозга?» (или «Является ли свойство испытывать боль в момент времени *t* состоянием мозга?») ². Нельзя должным образом рассмотреть этот вопрос, не высказавшись сначала об особых правилах, выработанных в ходе развития «аналитической философии»; правилах, которые вместо того, чтобы положить конец всем концептуальным заблуждениям, сами представляют собой значительное концептуальное затруднение. Эти правила, представленные, конечно, скорее в имплицитном, чем явном виде в творчестве большинства философов-аналитиков, таковы: (1) утверждение вида «*A* есть *B*» (например, «ощущение боли есть нахождение мозга в определенном состоянии»), *правильно* только в том случае, если оно вытекает, в некотором смысле, из значения терминов *A* и *B*; и (2) утверждение вида «*A* есть *B*», может быть философски *информативным* только в том случае, если оно в некотором смысле представляет собой редукцию (например, утверж-

* Данная статья больше известна под названием «Природа ментальных состояний» (The nature of mental states // Putnam H. Mind, Language and Reality, Philosophical Papers, Vol. 2, pp. 429—440.) Впервые текст был опубликован под названием «Психологические предикаты» (Psychological predicates // Capitan, Merrill (eds.). Art, Mind and Religion, 1967.

¹ Эти и связанные с ними вопросы я обсуждал в главах 16, 18 и 20 в кн.: Mind, Language and Reality, Philosophical Papers, Vol. 2. — Прим. Х. Патнэма.

² В этой статье я не буду касаться без конца обсуждаемого вопроса об отношении между *болью* и *состоянием боли*. Я лишь мимоходом отмечу, что можно легко показать ошибочность аргумента, выдвигаемого *против* их отождествления и состоящего в том, что боль может быть локализована в руке человека, а состояние (организма) не может быть локализовано в руке.

дение «ощущение боли есть неприятное ощущение» не является философски информативным, а утверждение «ощущение боли есть предрасположенность к определенному поведению» философски информативно, если оно истинно). Эти правила превосходны, если мы продолжаем верить в возможность осуществления программы редукционистского анализа (в стиле 1930-х годов); если же мы не верим в возможность такого анализа, то эти правила превращают аналитическую философию в занятие для дураков, по крайней мере, в том, что касается вопросов «что есть то-то и то-то».

В этой статье я буду использовать термин «свойство» в качестве общего термина для таких вещей, как испытывать боль, находиться в определенном состоянии мозга, иметь предрасположенность к определенному поведению, а также для таких величин, как температура и т. п., т. е. для вещей, которые могут быть естественным образом представлены одноместными или многоместными предикатами или функторами. Я буду использовать термин «понятие» для обозначения вещей, которые можно отождествить с классами синонимии (synonymy-class) соответствующих выражений. Так, понятие *температуры* может быть отождествлено (я считаю) с классом синонимии слова «температура»³. (Аналогичным образом мы говорим, что число 2 можно отождествить с классом всех пар. Но это существенно отличается от того странного утверждения, что 2 *есть* класс всех пар. Я не утверждаю, что понятия *есть* классы синонимии, чтобы это ни означало. Я считаю,

³ Хорошо известны замечания Алонзо Чёрча по этому вопросу. Они не имеют отношения (как может показаться на первый взгляд) к отождествлению понятий с классами синонимии как таковыми; скорее, они были высказаны в поддержку той точки зрения, что (в формальной семантике) необходимо сохранить введенное Фреге различия между прямым и «косвенным» употреблением выражений. То есть, даже если мы говорим, что понятие температуры *есть* класс синонимии слова «температура», то это не должно давать повод для ошибочного предположения, что «понятие температуры» синонимично «классу синонимии слова “температура”», поскольку тогда «понятие температуры» и «der Begriff der Temperatur» не были бы синонимами, хотя они таковыми являются. Скорее, мы должны сказать, что понятие «температура» *отсылает к* (refer to) классу синонимии слова «температура» (при этой конкретной интерпретации); но в этом случае данный класс *определяется* не как «класс синонимии, к которому принадлежит такое-то слово», а иным способом (например, как класс синонимии, элементы которого имеют такое-то характерное употребление).

то их можно отождествлять с классами синонимии в целях формализации соответствующих рассуждений).

Вопрос «Что такое понятие *температуры*?» звучит очень странно. Для кого-то он мог бы означать следующее: «Что такое температура? Пожалуйста, отнеситесь к моему вопросу как к концептуальному». В этом случае ответ мог бы быть таким (представим на мгновение, что слова «тепло» и «температура» — синонимы): «температура это тепло» или даже: «понятие температуры — то же самое понятие, что и понятие тепла». Для кого-то другой этот вопрос мог бы означать: «Что есть *понятия*? Например, то есть “понятие температуры”?». В этом случае ответ известен только богу. (Возможно, ответ состоял бы в том, что понятия *можно отождествить* с классами синонимии).

Конечно, вопрос «Что такое свойство *температура*?» тоже странен. Его можно истолковать как вопрос о понятии температуры. Однако физик не будет воспринимать его таким образом.

Когда мы утверждаем, что свойство P_1 может быть тождественным свойству P_2 только в том случае, если термины P_1 и P_2 в некотором соответствующем смысле «синонимичны», мы, по сути, вводим два понятия «свойство» и «понятие» — в одно. Точку зрения, согласно которой понятия (интенционалы) *есть* то же самое, то свойства, открыто отстаивал Карнап (например, в книге «*Знание и необходимость*»). Думаю, это неудачная точка зрения, поскольку утверждение «температура — это средняя молекулярно-кинетическая энергия» дает прекрасный пример истинного утверждения о тождестве свойств, в то время как утверждение «понятие температуры — это то же самое понятие, что и понятие средней молекулярно-кинетической энергии» попросту ложно.

Многие философы полагают, что утверждение «боль есть состояние мозга» нарушает некоторые правила или нормы естественного языка. Однако аргументы, выдвигаемые в пользу этой точки зрения, едва ли убедительны. Например, если на том основании, что я могу знать, что испытываю боль, не зная, что мой мозг находится в состоянии S , делается вывод, что боль не может быть состоянием мозга S , то таким же рассуждением можно доказать, что если я могу знать, что печь горяча, не зная, что велика ее средняя молекулярно-кинетическая энергия (или даже не зная, что существуют молекулы), то это опровергает утверждение, что температура *есть* средняя молекулярно-кинетическая энергия, а это уже противоречит физике. В действительности, на том основании, что я могу знать, что испытываю боль, не зная, что мой мозг находится в со-

стоянии S , можно непосредственно заключить только то, что понятие боли и понятие нахождения мозга в состоянии S — это разные понятия. Но это не мешает тому, чтобы или боль вообще, или состояние боли, или какая-то боль, или какое-то состояние боли были состояниями мозга S . В конце концов, понятие «температуры» не является тем же самым понятием, что и понятие средней молекулярно-кинетической энергии, однако температура — это средняя молекулярно-кинетическая энергия.

Некоторые философы считают оба утверждения «боль есть состояние мозга» и «состояния боли есть состояния мозга» неясными. В ответ нужно разъяснить этим философам — насколько это возможно, учитывая неопределенность всей научной методологии, — какого рода соображения заставляют нас осуществлять эмпирическую редукцию (т. е. утверждать такие вещи, как «вода есть H_2O », «свет есть электромагнитное излучение», «температура есть средняя молекулярно-кинетическая энергия»). Если, не приводя оснований, человек продолжает, невзирая на примеры, настаивать на том, что невозможно представить себе сходные обстоятельства в случае утверждений «боль есть состояние мозга» (или, возможно, «состояния боли есть состояния мозга»), то вполне оправданно считать его упрямым.

По мнению некоторых философов, если в предложении « P_1 есть P_2 » связка «есть» выражает эмпирическую редукцию, то это предложение может быть истинным только в том случае, если свойства P_1 и P_2 (а) связаны с некоторой пространственно-временной областью и (б) эта область одна и та же в обоих случаях. Так, утверждение «температура есть средняя молекулярно-кинетическая энергия» является допустимой эмпирической редукцией, поскольку температура и молекулярная энергия связаны с одной и той же пространственно-временной областью, а утверждение «ощущение боли в руке есть нахождение мозга в некотором состоянии» не является допустимой эмпирической редукцией, поскольку речь здесь идет о разных пространственных областях.

Этот аргумент не представляется очень веским. Тот факт, что отражение может быть «локализовано» в трех футах за зеркалом никому не мешает утверждать, что отражения в зеркале — это свет, отражаемый сначала объектом, а затем поверхностью зеркала! (Более того, если кто-то готов разрешить какие-то определенные редукции (как, например, температура есть средняя молекулярно-кинетическая энергия), то у этих редукций он всегда найдет некоторое общее свойство, отсутствующее у тех отождествлений

которые он хотел бы запретить. Это будет аргументом, если только удастся показать, что от наличия указанного общего свойства зависит, имеет ли смысл подобное отождествление или нет).

С другой стороны, ряд философов утверждает, что все предсказания, выводимые из конъюнкции нейрофизиологических законов и утверждений типа «состояния боли есть такие-то состояния мозга», можно вывести и из конъюнкции тех же самых нейрофизиологических законов и утверждений типа «ощущение боли имеет корреляцию с такими-то состояниями мозга», и, следовательно, (*sic!*), не может быть никаких методологических оснований, чтобы утверждать, что боль (или состояния боли) *есть* состояния мозга, вместо того, чтобы утверждать, что состояния боли имеют (инвариантную) корреляцию с состояниями мозга. Этот аргумент доказывал бы также, что между светом и электромагнитным излучением имеется лишь корреляция. Здесь игнорируется тот факт, что, хотя рассматриваемые теории действительно могут вести к одним и тем же предсказаниям, они ставят и исключают разные *вопросы*. Утверждение «Свет имеет инвариантную корреляцию с электромагнитным излучением» оставляет открытыми вопросы: «Что же тогда есть свет, если он не то же самое, что и электромагнитное излучение?» и «Почему свет сопровождает электромагнитное излучение?» — вопросы, которые исключаются, когда мы говорим, что свет *есть* электромагнитное излучение. Сходным образом, когда мы утверждаем, что боль есть состояние мозга, наша цель — показать, что вопросы: «Что же тогда есть боль, если она не то же самое, что и состояние мозга?» и «Почему боль сопровождает состояния мозга?» лишены эмпирической осмысленности (*meaningfulness*). Если есть основания предположить, что эти вопросы задают, так сказать, неверный подход к проблеме, то именно эти основания и оправдывают теоретическое отождествление боли и состояния мозга.

Если все контраргументы неубедительны, то означает ли это, что предложение «боль есть состояние мозга» или «состояния боли есть состояния мозга» осмысленны (и, возможно, истинны)?

(1) Предложение «боль есть состояние мозга» совершенно осмысленно (не нарушает «правил естественного языка», не предполагает «расширения употребления»).

(2) Предложение «боль есть состояние мозга» не является осмысленным (предполагает «изменение значения» или «расширение употребления» и т. п.).

Мою собственную позицию не выражает ни предложение (1), ни предложение (2). На мой взгляд, понятия «изменения значения» и «расширения употребления» столь нечетко определены, что фактически нельзя утверждать *ни* (1), *ни* (2). Я не вижу оснований считать, что лингвисты, простые люди или философы располагают на сегодняшний день понятием «изменения значения», применимым к случаям, которые мы здесь обсуждаем. Понятие изменения значения было введено в ходе развития языка для выполнения *намного* более простой *работы*, чем та, что возлагается на него в рассматриваемом случае.

Но, если мы не утверждаем ни (1), ни (2) — иными словами, если мы считаем вопрос об «изменении значения» псевдопроблемой в данном случае, то как нам следует разбирать вопрос, с которого мы начали, — «Является ли боль состоянием мозга?».

Наш ответ: допустив возможность утверждений вида «боль есть *A*», где «боль» и «*A*» никак не синонимы, попытаться установить, нельзя ли найти такое утверждение, для принятия которого имелись бы методологические и эмпирические основания. Именно этим мы сейчас и займемся.

II. ЯВЛЯЕТСЯ ЛИ БОЛЬ СОСТОЯНИЕМ МОЗГА?

Теперь мы рассмотрим вопрос «Является ли боль состоянием мозга?», оставив в стороне, как мы решили, проблему «изменения значения».

Поскольку я рассматриваю не вопрос, в чем состоит понятие боли, а вопрос, что есть боль, где связка «есть» предполагает построение эмпирической теории (или, по крайней мере, выдвижение эмпирического предположения), мне простительно будет предложить здесь некоторую эмпирическую гипотезу. По сути, моя стратегия обоснования того, что боль *не есть* состояние мозга, будет опираться не на априорные основания, а на тот факт, что другая гипотеза является более правдоподобной. Детальная разработка и верификация моей гипотезы была бы столь же утопичной задачей, как и подробная разработка и верификация гипотезы о боли как состоянии мозга. Однако выдвижение гипотез не в детально сформулированном и «законченном» виде, а в виде плана давно стало функцией философии. Короче говоря, я постараюсь обосновать, что боль — это не состояние мозга, если под этим понимать физико-химическое состояние мозга (или даже состояние всей нервной системы), а состояние совершенно иного *типа*. Я выдвигаю гипотезу

тезу, что боль или состояние боли — это функциональное состояние всего организма.

Для объяснения этой гипотезы необходимо ввести несколько специальных понятий. В предыдущих своих работах я разъяснил понятие *машины Тьюринга* и рассмотрел, как его можно использовать в качестве модели организма. Понятие *вероятностного автомата* определяется аналогично машине Тьюринга, за тем исключением, что здесь допускаются не столько «детерминистические» переходы между «состояниями», сколько переходы с различными степенями вероятности. (Машина Тьюринга — это просто особый вид вероятностного автомата, где вероятности перехода равны 0, 1). Я буду исходить из того, что в обобщенном виде понятие вероятностного автомата допускает «сенсорные входные данные» (sensory inputs) и «двигательные выходные данные» (motor outputs), т. е. *машинная таблица* для каждой возможной комбинации «состояния» и полного набора «сенсорных входных данных» содержит «команду», которая определяет вероятность следующего «состояния» и вероятности «двигательных выходных данных». (Это заменяет идею машины, печатающей на ленте.) Также я буду предполагать, что известна физическая реализация сенсорных органов, ответственных за различные входные данные, и двигательных органов, но сами «состояния» и «входные данные», как и обычно, определяются только «имплицитно», т. е. через множество вероятностей перехода, задаваемых машинной таблицей.

Поскольку реальная система одновременно может быть «физической реализацией» многих различных вероятностных автоматов, я введу понятие *описания* системы. Описанием системы S , где S — система, является любое истинное утверждение о том, что S имеет различные состояния S_1, S_2, \dots, S_n , такие, что их связи друг с другом, с двигательными выходными данными и с сенсорными входными данными определяются вероятностями перехода, задаваемыми такой-то машинной таблицей. Упомянутая в описании машинная таблица будет поэтому называться *функциональной организацией* системы S относительно данного описания, а состояние S_i такое, что S находится в S_i в некоторый данный момент времени, будет называться *совокупным состоянием* (total state) системы S (в данный момент времени) относительно данного описания. Нужно отметить, что знание совокупного состояния системы относительно данного описания включает множество сведений о том, как система склонна себя «вести» при различных комбинациях сенсорных входных данных, но не предполагает знания физи-

ческой реализации S_i , т. е. не предполагает, например, знания физико-химического состояния мозга. Повторяю, описание лишь *имплицитно* определяет S_i , т. е. определяет его *только* через множество вероятностей перехода, задаваемых машинной таблицей.

Теперь можно более точно сформулировать гипотезу о том, что «ощущение боли есть функциональное состояние организма»:

(1) Все организмы, способные ощущать боль, есть вероятностные автоматы.

(2) Любой организм, способный ощущать боль, имеет по крайней мере одно описание определенного вида (т. е. способность ощущать боль *есть* обладание функциональной организацией соответствующего вида).

(3) Ни один организм, способный ощущать боль, не предполагает разложения на части, которые по отдельности имеют описание, указанного в (2) вида.

(4) Для каждого указанного в пункте (2) описания существует подмножество сенсорных входных данных таких, что организм, имеющий такое описание, ощущает боль тогда и только тогда, когда некоторые из его сенсорных входных данных принадлежат к этому подмножеству.

Эта гипотеза, надо признаться, сформулирована нечетко, но столь же нечетко формулируется в настоящий момент и гипотеза о боли как состоянии мозга. Например, хотелось бы больше знать о том, каким видом функциональной организации должен обладать организм, способный ощущать боль, а также больше знать о том, какие признаки отличают подмножество сенсорных входных данных, упоминаемых в пункте (4). В ответ на первый вопрос, вероятно, можно указать, что функциональная организация должна включать что-то вроде «функции предпочтения» или хотя бы частичного упорядочения предпочтений и что-то вроде «индуктивной логики» (т. е. машина должна уметь «учиться на опыте»). (Значение этих условий для автоматических моделей обсуждалось в статье «*Ментальная жизнь некоторых машин*»⁴). Кроме того, вполне естественно предположить, что машина должна иметь «болевы датчики», т. е. сенсорные органы, которые, как правило, посылают сигналы о нанесенных телу машины повреждениях, об опасно высоких и низких температурах, давлениях и т. п. и которые передают особое подмножество входных данных, подмножест-

⁴ The mental life of some machines // Putnam H. Mind, Language and Reality. Philosophical Papers, Cambridge..., 1975, Vol. 2.

во, упоминаемое в пункте (4). Наконец, что касается второго вопроса, то, по крайней мере, необходимо, чтобы входным данным из выделенного подмножества функция предпочтения или упорядочение предпочтений предписывали высокое негативное значение (дополнительные условия также обсуждались в статье «*Ментальная жизнь некоторых машин*» (см. сноску 4)). Условие (3) предназначено для того, чтобы исключить из рассмотрения такие «организмы» (если их вообще можно считать организмами), как рой пчел. Условие (1), очевидно, излишне и введено только в целях разъяснения. (По сути оно пусто, поскольку при *определенном* описании любой объект является вероятностным автоматом).

Замечу мимоходом, что эта гипотеза, несмотря на ее неопределенность, гораздо *менее* неопределенна, чем современная гипотеза о «физико-химических состояниях», и представляет собой гораздо лучший объект для математического и эмпирического исследования. Действительно, для разработки этой гипотезы нужно просто попытаться создать «механические» модели организмов, но разве не это составляет, в некотором смысле, предмет психологии? Конечно, будет нелегко перейти от моделей к *конкретным* организмам и определить *нормальную форму* (normal form) для психологического описания организмов, ибо для этого потребуются точно сформулировать условие (2) и (4). Однако, думаю, психология и должна этим заниматься.

Теперь я сравню только что выдвинутую гипотезу с (а) гипотезой о том, что боль есть состояние мозга и (б) гипотезой о том, что боль есть предрасположение к определенному поведению.

III. ФУНКЦИОНАЛЬНОЕ СОСТОЯНИЕ ИЛИ СОСТОЯНИЕ МОЗГА

Быть может, я не совсем справедлив к разработчику теории состояний мозга, приписывая ему трактовку этих состояний как *физико-химических*? Но (а) физико-химические состояния мозга — это единственный вид состояний, когда-либо упоминавшихся разработчиками теории состояний мозга. (б) Разработчики теории состояний мозга обычно ссылаются (с гордостью, немного отдающей деревенским атеизмом) на несовместимость их гипотезы со всеми формами дуализма и ментализма, что вполне оправданно, если речь идет о физико-химических состояниях мозга. Однако функциональные состояния целых систем — нечто совсем иное. В частности, гипотеза о функциональных состояниях *совместима* с дуализмом! Хотя нет никаких сомнений, что эта гипотеза «механистична» по

своему духу, есть нечто примечательное в том, что система, состоящая из тела и «души», при условии, что душа существует, вполне может быть вероятностным автоматом. (в) Согласно аргументу, выдвинутому Смарттом, теория состояний мозга предполагает только «физические» свойства, а «не-физические» свойства Смартт считает недоступными пониманию (*unintelligible*). Совокупные состояния и «входные данные», как они определены выше, конечно же, не являются ни ментальными, ни физическими *per se*, и я не могу представить себе функционалиста, выдвигающего подобный аргумент. (г) Если разработчик теории состояний мозга имеет в виду (или, по крайней мере, допускает) иные, не физико-химические, состояния, то его гипотеза совершенно пуста, во всяком случае если он не уточнит, *какого рода «состояния» он имеет в виду.*

Если трактовать гипотезу о состояниях мозга в таком ключе, то какие есть основания считать, что гипотеза о функциональных состояниях более предпочтительна, чем гипотеза о состояниях мозга? Рассмотрим что требуется от разработчика теории состояний мозга, чтобы выдвигаемые им положения были убедительны. Он должен определить физико-химическое состояние таким образом, чтобы *любой* организм (а не только млекопитающее) только в том случае ощущал боль, если (а) у него есть мозг с соответствующей физико-химической структурой и если (б) его мозг находится в указанном физико-химическом состоянии. Это означает, что в данном физико-химическом состоянии может находиться мозг млекопитающего, мозг рептилии, мозг моллюска (осьминоги — это моллюски и, стало быть, ощущают боль) и т. д. С другой стороны, в этом состоянии *не может* (физически не может) находиться мозг ни одного физически возможного существа, неспособного ощущать боль. Если такое состояние удастся открыть, то должно быть номологически верно, что в этом состоянии будет находиться мозг любого обнаруженного нами неземного живого существа, которое способно ощущать боль, причем способно до того, как мы выдвинем предположение, что, *возможно*, — это боль.

Нет ничего невозможного в том, что такое состояние будет открыто. Хотя осьминоги и млекопитающие представляют собой пример параллельной (а не последовательной) эволюции, у тех и у других сформировались фактически идентичные (с физической точки зрения) структуры глаза, хотя данный орган развился в этих двух случаях из клеток различных типов. Поэтому, по крайней мере, есть такая возможность, что в ходе параллельной эволюции во

всей Вселенной *всякий раз* возникает *один и тот же* физический «коррелят» боли. Но это, безусловно, очень смелая гипотеза.

В конце отметим, что эта гипотеза окажется еще более смелой, когда мы осознаем, что разработчик теории состояний мозга не просто утверждает, что *боль* есть состояние мозга; он, конечно же, имеет в виду, что *каждое* психологическое состояние есть состояние мозга. Поэтому, если мы сможем найти хотя бы один психологический предикат, применимый и к млекопитающим, и к осьминогам (скажем, предикат «голодный»), но имеющий разные физико-химические «корреляты», то теория состояний мозга будет опровергнута. Вероятность того, что мы сможем найти такой предикат, представляется мне очень высокой. Безусловно, в этом случае разработчик теории состояний мозга может прибегнуть к допущениям *ad hoc* (например, он сможет определить дизъюнкцию двух состояний как единое «физико-химическое состояние»), но к этому не стоит относиться всерьез.

Обращаясь теперь к соображениям *в пользу* теории функциональных состояний, отметим прежде всего, что мы приписываем организмам ощущения боли, голода, гнева и т. п., на основании их *поведения*. Общеизвестно, что сходство в поведении двух систем дает, по крайней мере, *больше* оснований, чтобы предположить сходство в их функциональной организации, чем сходство в их реальных физических деталях. Более того, мы рассчитываем на то, что разные психологические состояния, или хотя бы, основные из них, такие, как голод, жажда, агрессия и т. п., имеют более или менее сходные «вероятности перехода» (разумеется, в нечетко определенных и довольно широких пределах), если сравнивать поведение животных разных биологических видов, поскольку именно по этим признакам мы распознаем эти состояния. Так, мы не будем считать, что животное *испытывает жажду*, если его поведение не будет направлено на поиск питья и не завершиться состоянием «насыщения жидкостью». Поэтому любое животное, которое мы считаем способным находиться в этих разнообразных состояниях, по крайней мере, *будет выглядеть так*, будто обладает некоторой функциональной организацией. Как уже отмечалось, если когда-нибудь нам удастся открыть безотносительные к биологическому виду психологические законы, например, удастся определить нормальную форму для психологических теорий, то в результате мы сможем описать виды функциональной организации, необходимые и достаточные для того или иного психологического состояния, а также точно определить понятие «психологического состояния».

Напротив, разработчику теории состояний мозга приходится уповать на то, что в конце концов будут разработаны безотносительные к биологическому виду нейрофизиологические законы, но, думаю, есть больше оснований надеяться на то, что можно будет сформулировать безотносительные к биологическому виду психологические законы (достаточно общего вида) или хотя бы можно будет найти безотносительную к биологическому виду *форму* записи психологических законов.

V. ФУНКЦИОНАЛЬНОЕ СОСТОЯНИЕ ИЛИ ПРЕДРАСПОЛОЖЕННОСТЬ К НЕКОТОРОМУ ПОВЕДЕНИЮ

Теория, согласно которой ощущение боли — это не состояние мозга и не функциональное состояние, а предрасположенность к определенному поведению, имеет одно явное преимущество: она, похоже, согласуется с тем, каким образом мы устанавливаем, что организм ощущает боль. Когда мы говорим, что животное ощущает боль, мы по сути ничего не знаем о состоянии его мозга и у нас почти нет никаких представлений о его функциональной организации, за исключением, пожалуй, самых приблизительных и интуитивных. Однако упомянутое «преимущество» вовсе и не преимущество, ибо хотя, возможно, наш способ установления того, что *x* есть *A*, имеет отношение к нашему понятию об *A*, но он почти ничего не говорит о том, что *есть* свойство *A*. Если на основе упомянутого «преимущества» мы отрицаем, что боль — это состояние мозга или функциональное состояние, то это равносильно тому, как если бы мы отрицали, что тепло есть средняя молекулярно-кинетическая энергия, на том основании, что простые люди не определяют (по их мнению) среднюю молекулярно-кинетическую энергию предмета, когда устанавливают горячий он или холодный. Они и не должны этого делать; необходимо, чтобы средняя молекулярно-кинетическая энергия позволяла объяснить то, что они считают признаками тепла. Аналогичным образом, в случае нашей гипотезы необходимо, чтобы поведенческие признаки боли объяснялись тем фактом, что организм находится в функциональном состоянии соответствующего вида, но совершенно необязательно, чтобы говорящие *знали* об этом.

Трудности, связанные с описанием «предрасположенности к определенному поведению» столь хорошо известны, что я лишь перечислю их здесь. Главная трудность (на самом деле, думаю, это

больше, чем просто «трудность») состоит в том, что нужную в данном случае предрасположенность нельзя определить иначе, как «предрасположенность X вести себя так, как если бы X ощущал боль». Мы, напротив, *можем*, не используя понятия «боль», определить, по крайней мере, приблизительно то функциональное состояние, с которым предлагаем отождествить боль. Под этим функциональным состоянием мы имеем в виду состояние получения организмом сенсорных входных данных, играющих определенную роль в функциональной организации этого организма. Эту роль, по крайней мере, частично характеризует тот факт, что функция органов чувств, ответственных за указанные входные данные, состоит в выявлении нанесенных телу повреждений, опасно высоких и низких температур, давлений и т. п. Эту роль характеризует и то, что сами «входные данные», безотносительно к их физической реализации, представляют собой состояние, которому организм приписывает крайне негативное значение. Но это *не* означает, что машина всегда будет *избегать* находиться в указанном состоянии («состоянии боли»), как я подчеркивал в другой своей статье (см. сноску 4). Это означает лишь, что она будет избегать находиться в этом состоянии, если только нахождение в этом состоянии не будет необходимо для достижения какой-то иной, более важной цели. Поскольку поведение организма будет определяться не только сенсорными входными данными, но и его совокупным состоянием (т. е. другими ценностями, убеждениями и т. п.), то, видимо, безнадежно пытаться высказывать общие суждения о том, как *должен* вести себя организм в подобных условиях. Но это не означает, что мы должны оставить надежду охарактеризовать эти условия. По сути дела, мы только что их охарактеризовали ⁶.

Но теория «предрасположенности к определенному поведению», видимо, не только безнадежно неопределенна. Если «поведение», о котором идет речь, является внешним (peripheral) поведением, а соответствующие стимулы — внешними стимулами (напри-

⁶ В статье «Ментальная жизнь некоторых машин» описывается в терминах автоматических моделей еще одна, отчасти независимая характеристика болевых входных данных, а именно: спонтанная склонность отдергивать поврежденную часть и т. п. Это ставит вопрос о функциональном анализе понятия спонтанной склонности, который и обсуждается в указанной статье. Конечно, на ум сразу приходят и другие характеристики, например, *локализация* (действительная или кажущаяся) ощущения боли в частях тела.

мер, мы не рассматриваем, что будет делать организм, если его мозг оперируют), то эта теория просто ложна. Так, два животных, которым удалили двигательные нервы, будут потенциально и реально «демонстрировать» одинаковое поведение (то есть, его отсутствие). Если при этом одному из них удалить болевые окончания, а другому — нет, то одно животное будет ощущать боль, а другое — нет. С другой стороны, если один человек не имеет болевых окончаний, а другой сознательно, огромным усилием воли не позволяет себе реагировать на боль, то их поведение может быть совершенно одинаковым, но один будет ощущать боль, а другой — нет. (Некоторые философы считают этот последний пример концептуально невозможным, но, похоже, единственным аргументом служит их неспособность или нежелание представить себе подобный случай ⁶.) Если вместо боли мы возьмем ощущение, телесное проявление которого легче сдерживать, например, небольшое ощущение холода в левом мизинце, то пример станет еще более очевидным.

В конце концов, даже если бы *существовала* предрасположенность к определенному поведению, которая имела бы инвариантную (и безотносительную к биологическому виду!) корреляцию с болью и которую можно было бы определить без помощи термина «боль», все же было бы более правдоподобно (plausible) отождествить ощущение боли не с самой этой предрасположенностью к определенному поведению, а с некоторым состоянием — состоянием мозга или функциональным состоянием — которое *объясняет* эту предрасположенность. Возможно, соображения правдоподобия несколько субъективны, но будь прочие условия равны (разумеется, они не равны), почему бы не позволить соображениям правдоподобия сыграть решающую роль?

V. МЕТОДОЛОГИЧЕСКИЕ СООБРАЖЕНИЯ

До сих пор мы рассматривали гипотезу о том, что ощущение боли есть функциональное состояние, а не состояние мозга или предрасположенность к определенному поведению, лишь в свете, так сказать, ее «эмпирических» оснований, то есть мы учитывали лишь бóльшую вероятность предположения об инвариантной и безотносительной к биологическому виду «корреляции» между болью и функциональным состоянием по сравнению с предположени-

⁶ Ср. обсуждение «супер-спартанцев» в гл. 16 кн.: Putnam H. *Mind, Language and Reality*. Philosophical Papers, Cambridge..., 1975, Vol. 2.

ем о подобной корреляции между болью и физико-химическим состоянием мозга (должен ли организм иметь мозг, чтобы ощущать боль?; быть может достаточно нервных узлов?) или между болью и предрасположенностью к определенному поведению. Если это так, то отсюда следует, что предложенное нами отождествление, по крайней мере, достойно обсуждения. Но что можно сказать о методологических соображениях?

Во всех случаях редукции методологические соображения приблизительно одинаковы, поэтому здесь не стоит ждать неожиданностей. Во-первых, отождествление психологических состояний с функциональными означает, что из утверждений вида «такие-то организмы имеют такие-то описания» и утверждений тождества («ощущение боли есть нахождение в таком-то функциональном состоянии» и т. п.) можно вывести законы психологии. Во-вторых, функциональное состояние (т. е. входные данные, играющие в функциональной организации организма описанную выше роль) не просто «имеет корреляцию с» поведением организма, ощущающего боль, но и объясняет это поведение. В-третьих, отождествление позволяет исключить вопросы, представляющие (с натуралистической позиции) совершенно неверный подход к решению проблемы. Сюда относятся вопросы: «что же *есть* боль, если она не состояние мозга и не функциональное состояние?» или «почему боль всегда сопровождается таким-то функциональным состоянием?». Коротче говоря, это отождествление нужно предварительно принять как теорию, которая ведет к плодотворным предсказаниям и к постановке плодотворных *вопросов*, а также позволяет устранить пустые и эмпирически бессмысленные вопросы. Под «эмпирически бессмысленными» я имею в виду вопросы, «бессмысленные» не только с точки зрения верификации, но и с точки зрения того, что *существует* в реальности.

1.

В настоящей статье я хотел бы рассмотреть характер различных «менталистских» понятий на основе машинного аналога. В статье «Сознание и машины»¹ я попытался показать, что концептуальные вопросы, порождаемые традиционной психофизической проблемой (*mind—body problem*), никак не связаны с предположительно особым характером человеческого субъективного опыта; они возникают в связи с любой компьютерной системой определенного уровня сложности, в частности, в связи с любой компьютерной системой, способной строить теории о своей собственной природе. В упомянутой статье меня интересовали прежде всего вопросы, касающиеся тождества сознания и мозга. В настоящей статье главное внимание будет уделено прояснению характера понятий, выражаемых такими словами, как «предпочитать», «считать» (*believe*), «чувствовать». Рассматривая, как употребляются эти слова применительно к машинному аналогу, я надеюсь показать, что традиционные альтернативы — материализм, дуализм и логический бихевиоризм — неверны даже в случае машин. Но мои цели — не только деструктивные; я надеюсь — установив характер этих слов в случае машинного аналога — выдвинуть некоторые предположения о том, каков их характер, когда они применяются к людям.

Я не буду обсуждать — если не считать этих сделанных вначале замечаний, — вопрос о том, в какой мере применение таких терминов, как «предпочтение», к машинам Тьюринга представляет собой изменение или расширение значения этого термина. Я не буду обсуждать этот вопрос, ибо в дальнейшем станет ясно, что он не имеет прямого отношения к моему исследованию. Хотя когда мы говорим об описываемых мной машинах Тьюринга, что они «предпочитают» одну вещь другой, то, что мы имеем в виду, во *многих* отношениях *сильно* отличается от того, в каком смысле мы говорим о человеке, что он предпочитает одну вещь другой, это не противоречит ни одному из выдвигаемых мной утверждений. Я утверждаю, что, поняв, почему аналоги материализма, дуализма и логического бихевио-

* «The mental life of some machines». Впервые работа опубликована в кн.: *Castaneda H. (ed.) Intentionality, Minds and Perception*, Detroit, 1967.

¹ *Putnam H. Mind and machines // Mind, Language and Reality, Philosophical Papers. Vol. 2. Cambridge: Cambridge University Press, 1975, p. 362—385.*

ризма ложны в случае машин Тьюринга, мы сможем понять, почему упомянутые теории неверны и в случае их применения к людям, а установив, что эти термины могут означать в случае машин Тьюринга, мы сможем усмотреть важные логические характеристики этих терминов, которые прежде не были отмечены философами.

Итак, в этой статье я рассмотрю гипотетическое «сообщество» «агентов», каждый из которых является машиной Тьюринга или, если говорить точнее, конечным автоматом. (Из множества приемлемых эквивалентных определений «конечного автомата» для наших целей лучше всего подходит то, которое мы получим из определения машины Тьюринга, уточнив, что лента для печати должна быть *конечной*.) Машины Тьюринга, которые я предполагаю рассмотреть, отличаются от абстрактных машин Тьюринга, описываемых логической теорией, тем, что они оснащены органами чувств, с помощью которых они могут изучать окружающую среду, и имеют соответствующие двигательные органы, которыми они могут управлять. Можно считать, что органы чувств обуславливают печать определенных «сообщений» на ленте машины в определенные моменты времени, а машина сконструирована таким образом, что когда определенные символы «команд» печатаются в определенные моменты времени на ленте машины, ее двигательные органы выполняют соответствующие действия. Благодаря такому естественному расширению машины Тьюринга возможно ее взаимодействие с окружающей средой.

В основном мы хотели бы обсудить понятие *предпочтения*. Чтобы формально описать это понятие относительно поведения наших «агентов», предположим, что каждому из них приписывается функция рационального предпочтения в том смысле, как ее определяет экономическая теория². Предположим, что наши машины Тьюринга достаточно сложны и могут более или менее правильно оценивать вероятность различных положений дел. С учетом сделанных машиной индуктивных оценок, ее поведение будет поэтому полностью определяться тем фактом, что она подчиняется правилу: действовать так, чтобы максимизировать предполагаемую полезность.

² Neumann J. von, Morgenstern O. A Theory of Games and Economic Behavior. 3rd ed. Princeton, N. J., 1953, p. 26f, 83 et al. Фон Нейман и Morgenstern трактуют эту функцию как приписывание координат (в n -мерном пространстве) объектам, при этом сумма координат выражает «ценность» объекта. В нашем случае удобней рассматривать эту функцию как приписывание «полезности» «возможным мирам» (или «описаниям состояний» в трактовке Карнапа).

Читателю следует заметить, что термин «полезность» здесь полностью элиминируем. Мы утверждаем лишь, что с каждой машиной связывается определенная математическая функция, называемая функцией полезности, такая, что вместе с другой функцией — «степенью подтверждения» — она полностью определяет поведение машины в соответствии с заданным правилом и заданными теоремами исчисления вероятностей³. Короче говоря, наши машины являются рациональными агентами в том смысле, в каком это понятие используется в индуктивной логике и в экономической теории. Если функции рационального предпочтения этих машин сходны с функциями рационального предпочтения идеализированных людей, а вычислительные способности машин приблизительно совпадают с вычислительными способностями людей, то поведение этих машин очень похоже на поведение (идеализированных) людей. Эту модель можно усложнить, внося в поведение машин некоторые элементы иррациональности, присутствующие в поведении реальных людей (например, элемент иррациональности, связанный с нарушением транзитивности предпочтений), но здесь мы не будем этого делать.

Что же означает слово «предпочитать» применительно к одной из этих машин? Начнем с того, что оно просто означает, что функция, которая управляет поведением машины (точнее говоря, функция, которая вместе с индуктивной логикой управляет поведением машины), приписывает первой альтернативе большую ценность, чем второй. Уже с самого начала мы можем понять, что связь между предпочтением и поведением будет у этих машин довольно сложной. Например, если одна из машин предпочитает альтернативу *A* альтернативе *B*, то отсюда не следует с необходимостью, что в любой конкретной ситуации она выберет *A*, а не *B*. Делая выбор между *A* и *B*, машина должна учесть, какие последствия будет, скорее всего, иметь этот выбор в данной конкретной ситуации и, возможно, машина примет в расчет иные «ценности», чем предпочтительность *A* по отношению к *B*. Мы могли бы сказать, что если машина предпочитает альтернативу *A* альтернативе *B*, то это означает, что *ceteris paribus*⁴ машина выберет *A*, а не *B*, но мы отчаялись бы когда-либо точно сформулировать это *ceteris paribus*. Аналогичным образом,

³ Carnap R. Logical Foundations of Probability. Chicago, 1950, в частности с. 253—70.

⁴ При прочих равных условиях (лат.) — *Прим. перев.*

мисс Энском ⁶ полагает, что если человек намерен избежать аварии, то это означает, что *ceteris paribus* он выберет такой способ вождения машины, который, скорее всего, минимизирует возможность аварии. По ее мнению, в такого рода случаях невозможно в принципе подробно сформулировать это *ceteris paribus*. На этом основании она выдвигает предположение о фундаментальном различии между — как она говорит — практическим разумом и научным разумом. Однако к этому выводу следует отнестись скептически. Фактически мисс Энском показала, что некоторые предложенные методы определения рассматриваемого *ceteris paribus* не работают, но эти методы не работали бы и в случае машин. Отсюда вряд ли следует, что наши машины в своем обычном «поведении» демонстрируют вид рассуждения, фундаментально отличный от научного рассуждения. Напротив, при заданной функции рационального предпочтения действовать всегда так, чтобы максимизировать предполагаемую полезность, значит демонстрировать научное мышление весьма высокого уровня.

Мисс Энском могла бы в ответ заметить, что реальные люди не имеют функций рационального предпочтения. Однако фон Нейман и Моргенштерн показали, — и их доказательство является важнейшим результатом в данной области, — что любой агент, предпочтения которого непротиворечивы, всегда ведет себя так, что его поведение можно истолковать в терминах по крайней мере одной функции рационального предпочтения. Мисс Энском могла бы возразить, что реальные люди не имеют непротиворечивых предпочтений; но это означало бы, что различие между практическим и научным разумом состоит в том, что практический разум, по сути, часто бывает более или менее иррациональным, и, что практические рассуждения любого человека иррациональны в каких-то областях. Это равносильно утверждению, что дедуктивная логика в принципе отличается от логики, излагаемой в учебниках, поскольку дедуктивные рассуждения любого человека неправильны в некоторых областях. Дело в том, что замечания мисс Энском об интенциях должны применяться не только к интенциональному поведению более или менее иррациональных людей, но и к интенциональному поведению идеального рационального человека, обладающего богатой и сложной системой ценностей. Думаю, это ясно следует из ее книги. Но в применении к такому агенту один из ее основных выводов просто ложен: практи-

⁶ *Anscombe G. E. M. Intention. Ithaca, N. Y., 1953, в частности с. 59—61.* Я хотел бы подчеркнуть, что критикуемая мною точка зрения — это лишь три страницы из книги, которую я считаю великолепной.

ческие рассуждения такого агента, как мы видели, совсем не отличаются от научных рассуждений⁶.

Короче говоря, практическое рассуждение *имеет* фундаментальное отличие от научного, если принять, что научное рассуждение состоит из силлогизмов, посылки которых в принципе допускают точную формулировку, а практическое рассуждение состоит из так называемых «практических силлогизмов», посылки которых должны во всех представляющих интерес случаях содержать неизлиминируемое *ceteris paribus*. Однако реальное научное рассуждение включает гораздо более сложные способы связи посылок и заключений, чем силлогизм, а принятие решений как в реальном, так и идеальном вариантах включает в себя виды рассуждений, которые получают крайне неточное описание, если их пытаются подогнать под традиционную модель «практического силлогизма». Сложный процесс взвешивания разнообразных и противоположных альтернатив и ценностей предполагает дедуктивную систематизацию, но иного ти-

⁶ Некоторые из отмеченных мисс Энском различий между практическим и теоретическим рассуждениями действительно имеют место. Например, верно, что в большей посылке [практического рассуждения] должно говориться о чем-то нужном, а заключением должно быть действие (хотя «ничто не мешает придумать словесную форму, которой он (человек) *сопровождает* свое действие, и тогда это действие можно назвать заключением в словесной форме» — *Ibid.*, р. 60). У меня вызывает возражение лишь утверждение, что заключение (в «словесной форме») не следует *дедуктивно* из посылок (по крайней мере, во многих случаях — ср. ее прим. 1: *Ibid.*, р. 58) и не должно следовать, если только большая посылка не является «абсурдной» (*insane*) и ни для кого неприемлемой. Это подводит мисс Энском к той точке зрения, что Аристотель по сути занимался «описанием порядка (*order*), который имеет место, когда действия совершаются интенционально» (стр. 79). Отсюда опасно близко до предположения, что практическое рассуждение — это просто совершение интенциональных действий. Мэри Мазерсил (в кн.: *Mothersill M. Anscombe's account of the practical syllogism // Philosophical Review*, LXXI, 1962, р. 448—461) критикует мисс Энском за то же самое, но, кажется, не учитывает силу ее аргумента. Мы можем утверждать, как это делает Мазерсил, что предположение «делай все, что способствует избежанию аварии» имеет «неабсурдную» интерпретацию; это, конечно, верно, но ничего нам не дает, поскольку *при «неабсурдной» интерпретации* предположение «делай это» не следует дедуктивно из большей посылки и утверждения «это способствует избежанию аварии», т. к. *это* может не быть соответствующим действием, а «делай все» означает (при неабсурдной интерпретации) «делай все соответствующее». Мазерсил, видимо, исходит из того, что можно сформулировать «условия, при которых что-либо предполагается соответствующим», но именно это и отрицает Энском.

па, чем дедуктивная систематизация, рассматриваемая мисс Энском (и Аристотелем).

Прежде чем продолжить наше рассмотрение, я хотел бы сделать одно замечание, которое, возможно, предупредит некоторые неправильные истолкования. Машина Тьюринга — это просто система, имеющая дискретное множество состояний, связанных определенными способами. Обычно считается, что машина Тьюринга обладает памятью в виде бумажной ленты, на которой она печатает символы; впрочем, это можно считать просто метафорой. В случае конечного автомата, т. е. машины Тьюринга, имеющей конечную, а не потенциально бесконечную ленту, физическую реализацию этой ленты можно представить в виде любой конечной системы запоминания информации. Под «символом» имеется в виду любого рода *знак* (trace), который можно поместить в запоминающее устройство, а затем «считать» с помощью того или иного механизма. Мы можем расширить наше определение, допустив, что «машина» «печатает» и сканирует более одного символа за раз. Тьюринг показал, что эти расширения по сути оставляют класс машин Тьюринга в неизменном виде. Заметьте, что машина Тьюринга необязательно должна быть *машиной*. Машина Тьюринга вполне могла бы быть и биологическим организмом. Является ли реальный человек или мозг человека машиной Тьюринга (или, скорее, конечным автоматом) — это эмпирический вопрос. Сегодня мы не знаем ничего, что противоречило бы, в строгом смысле, гипотезе о том, что все мы без исключения являемся машинами Тьюринга, хотя нам известно и то, что делает эту гипотезу маловероятной. Строго говоря, машина Тьюринга не должна быть даже физической системой: машиной Тьюринга может быть все, что с течением времени может проходить через последовательность состояний. Так, картезианцу-дуалисту, склонному считать человеческое сознание самодостаточной системой, имеющей некоторый вид каузального взаимодействия с телом, можно указать, что с точки зрения чистой логики человеческое сознание вполне может быть машиной Тьюринга (если допустить, что человеческое сознание обладает достаточно большим, но не бесконечным, множеством состояний, что, кажется, не вызывает сомнений). Человеку, полагающему, что у людей есть душа и в этой душе заключены личность и память, которые сохраняются после смерти тела, опять-таки можно сказать, что с точки зрения чистой логики человеческая душа вполне может быть машиной Тьюринга или, скорее, конечным автоматом.

Хотя вполне вероятно, что и состояния человеческого мозга, и человеческие ментальные состояния, образуют дискретные множест-

ва, — и здесь неважно, какое значение мы приписываем несколько неопределенному понятию ментального состояния, — однако маловероятно, чтобы сознание или мозг были машиной Тьюринга. Рассуждая *a priori*, мы сочли бы более правдоподобной ту точку зрения, что взаимосвязи между различными состояниями мозга и ментальными состояниями человека носят скорее вероятностный, а не детерминистский характер, и что важное значение имеет запаздывание во времени. Однако эмпирических свидетельств тому ничтожно мало. Причина кроется в том, что автомат, у которого связи между состояниями описываются вероятностными законами, а поведение предполагает запаздывание во времени, вполне можно смоделировать на основе поведения произвольно взятой машины Тьюринга. Поэтому в силу специфики рассматриваемого случая одних лишь эмпирических данных недостаточно, чтобы сделать выбор между гипотезой, утверждающей, что человеческий мозг (и, соответственно, сознание) — это машина Тьюринга, и гипотезой, утверждающей, что мозг представляет собой более сложный вид автомата с вероятностными связями и запаздыванием во времени.

Впрочем, еще в одном отношении наша модель, безусловно, является упрощением, даже если признать, что человеческий мозг и сознание *представляют собой* машины Тьюринга. Как уже отмечалось, необходимым и достаточным условием согласования поведения человека в некоторый момент времени с приписываемой ему функцией рационального предпочтения является непротиворечивость делаемых этим человеком выборов, т. е. если человек предпочитает альтернативу *A* альтернативе *B*, а альтернативу *B* альтернативе *C*, то он предпочитает альтернативу *A* альтернативе *C*. Но даже эта очень слабая аксиома транзитивности нарушается когда мы имеем дело с предпочтениями многих, а, возможно, и всех, реальных людей. Поэтому, вряд ли какая-либо схема предпочтений реального человека согласуется с приписываемой ему функцией рационального предпочтения. Более того, даже если чья-то схема предпочтений согласуется с приписываемой функцией рационального предпочтения, сомнительно, чтобы человек последовательно подчинялся правилу, предписывающему максимизировать предполагаемую полезность.

И, наконец, наша модель не динамична. Иными словами, она не допускает изменения функции рационального предпочтения во времени, хотя это можно исправить. Таким образом, наша модель в ряде аспектов сверхпроста и сверхрационалистична. Однако было бы нетрудно — в принципе, ибо на практике это может оказаться недостижимым — усложнить нашу модель во всех этих аспектах: сде-

дать ее динамичной, допустить элементы иррациональности в предпочтениях и индуктивной логике машины, допустить отклонения от правила, предписывающего максимизировать предполагаемую полезность. Но не думаю, что какое-либо из этих усложнений изменило бы философские выводы, сделанные в настоящей статье. Иными словами, я не думаю, что философские выводы настоящей статьи стали бы другими, если бы мы заменили понятие машины Тьюринга понятием *K*-машины, достаточно богатой и сложной, чтобы можно было бы утверждать о человеческом мозге и сознании, что они в буквальном смысле есть *K*-машины.

Я не буду ничего говорить о своих гипотетических «агентах» помимо того, что они являются машинами Тьюринга и имеют функции рационального предпочтения. Этими агентами могли бы быть искусственно созданные объекты, биологические организмы и даже люди. Именно поэтому нигде в своей статье я не уточняю, являются ли «агенты» в моем «сообществе» живыми существами, обладают ли они сознанием или нет. Впрочем, в одном смысле об этих агентах можно сказать, что они, независимо от их физической реализации, *осознают* одни вещи и *не осознают* другие. Более того, если у них есть периоды, соответствующие сну, то мы можем сказать, что они «осознают» что-то в одни периоды времени и «не осознают» — в другие.

2. МАТЕРИАЛИЗМ

Думаю, не нужно доказывать, что картезианский дуализм не постоятелен в качестве описания «внутренней жизни» рассматриваемых машин и связей между этой внутренней жизнью и их поведением. «Агенты» — это просто определенные системы состояний, имеющие определенные каузальные взаимосвязи; *все* эти состояния каузально взаимосвязаны. Не существует двух отдельных «миров» — «мира» «внутренних» состояний и «мира» «внешних» состояний, с каким-то особым видом связи или корреляции. «Агенты» — это не души, обитающие в машинах Тьюринга, а *сами* машины Тьюринга.

А как быть с материализмом? Если материализм как философская доктрина правильно описывает ментальную жизнь *любого* организма, то он *безусловно* должен быть правильным описанием и «ментальной жизни» *наших* агентов; по крайней мере, если мы представляем их себе в виде автоматов, составленных из триггеров, реле, вакуумных трубок и т. п. Но я постараюсь обосновать, что даже в этом последнем случае традиционный материализм неверен.

Традиционный материализм (во многом ставший сегодня для философов воображаемым противником, которого легко опроверг-

нуть) утверждает, что слова, описывающие ментальное поведение, определены с помощью понятий, выражающих физико-химическое строение. Если это так, то предикат « T предпочитает альтернативу A альтернативе B » должен быть определен в терминах физико-химического строения наших машин Тьюринга. Однако, если нам дана посылка, что одна из наших машин Тьюринга имеет такое-то физико-химическое строение, мы не можем сделать логически обоснованный вывод о том, что эта машина предпочитает альтернативу A альтернативе B в том смысле, как это было объяснено выше, а если нам дана посылка, что машина предпочитает альтернативу A альтернативе B , мы не можем сделать логически обоснованный вывод о том, что она имеет такое-то физико-химическое строение. Это логически независимые утверждения даже в том случае, если наши машины Тьюринга — это *только* машины.

В этом легко удостовериться. Допустим, в качестве посылки мы имеем, что T_1 предпочитает альтернативу A альтернативе B . Отсюда мы можем заключить, что T_1 должна быть определенным образом запрограммирована. В частности, ее программа должна содержать функцию рационального предпочтения, приписывающую A более высокую ценность, чем B . Предположим, что мы располагаем не только этой информацией, но и конкретной машинной таблицей машины T_1 . Все равно мы не сможем сделать никакого вывода о физико-химическом составе T_1 по той причине, что *одна и та же* машина Тьюринга (с учетом ее машинной таблицы) может быть физически реализована потенциально бесконечным числом способов. Даже если машина, принадлежащая к сообществу наших агентов, предпочитает альтернативу A альтернативе B тогда и только тогда, когда взведен триггер 57, это совершенно случайный факт. Наша машина могла бы быть точно такой же во всех своих «психологических» аспектах, не содержи она никаких триггеров.

А возможны ли выводы в обратном направлении? Предположим, у нас есть сведения, что машина T_1 имеет определенное физико-химическое строение, можем ли мы сделать отсюда вывод, что она имеет определенную функцию рационального предпочтения? Это сводится к вопросу: выводима ли машинная таблица из физико-химического строения машины? Несомненно, что в эмпирическом плане мы *можем* сделать такой вывод, по крайней мере в простых случаях. Но мы здесь рассматриваем вопрос о логически обоснованных, а не эмпирически успешных выводах. Чтобы знать, что машина имеет такую-то машинную таблицу, мы должны знать, сколько значимо различных состояний она способна иметь, и как они каузально

между собой связаны. А это нельзя вывести из физико-химического строения машины, если только в дополнение к знанию физико-химического строения мы не знаем определенных *законов природы*. Нам необязательно знать все законы природы; мы должны знать лишь необходимое и конечное их множество. Однако у нас нет способа определить, какое именно конечное множество законов природы нужно добавить к знанию физико-химического строения машины, чтобы можно было доказать, что данная машина имеет определенную машинную таблицу. На основе одного только факта наличия у машины определенного физико-химического строения нельзя сделать ни вывода о том, что она обладает, ни вывода о том, что она не обладает какой-то конкретной функцией рационального предпочтения, а, стало быть, нельзя сделать ни вывода о том, что она предпочитает, ни вывода о том, что она не предпочитает альтернативу *A* альтернативе *B*.

Если дано описание физико-химического строения машины и сформулированы все законы природы (для простоты предположим, что их конечное множество), можем ли мы заключить, что машина предпочтет альтернативу *A* альтернативе *B*? Предположим, определенности ради, что законы природы — это законы классической атомной физики; т. е. они описывают поведение отдельных элементарных частиц, и имеется композиционная функция, позволяющая определить поведение любого отдельно взятого комплекса элементарных частиц. В итоге, физико-химическое строение машины описывается определенным комплексом элементарных частиц. С точки зрения *чистой логики* даже в этом случае мы не можем на основе данных утверждений сделать вывод о том, что машина имеет такую-то машинную таблицу или такую-то функцию рационального предпочтения, если только помимо физико-химического строения машины и законов природы, не будет дано в виде дополнительной посылки (с формальной точки зрения логически независимое) утверждение о том, что мы располагаем описанием *всей* машины. Представим себе в качестве примера, что помимо элементарных частиц существуют «сгустки эктоплазмы», — в сущности, неизвестные физической теории, и что в целом машина состоит из элементарных частиц и нескольких «сгустков эктоплазмы», связанных между собой сложными каузальными отношениями. Поэтому когда мы указываем физико-химическое строение машины в его обычном понимании, мы описываем лишь часть *структуры* всей машины. На основе этого описания части структуры и законов природы в их обычном понимании (законов, управляющих поведением *отдельно взятых систем*

элементарных частиц) мы можем сделать вывод о том, как будет вести себя эта часть структуры *при условии, что она не взаимодействует с остальной частью структуры* («сгустками эктоплазмы»). Поскольку чистая логика не утверждает, что физико-химическое описание является описанием всей машины, то из описания физико-химического строения машины и законов природы нельзя средствами чистой логики вывести, что машина имеет такую-то конкретную машинную таблицу или такую-то конкретную функцию рационального предпочтения.

В логическом плане эта ситуация аналогична той, когда некоторые философы пытаются трактовать универсальные обобщения как (возможно бесконечные) конъюнкции, т. е. когда предлагается рассматривать обобщение «все вороны черные» как конъюнкцию «(a_1 есть ворон $\supset a_1$ черный) & (a_2 есть ворон $\supset a_2$ черный) & (a_3 есть ворон $\supset a_3$ черный)...», где a_1, a_2, \dots — возможно бесконечный список индивидуальных констант, обозначающих всех воронов. Ошибка здесь состоит в том, что хотя эта конъюнкция действительно вытекает из утверждения, что все вороны черные, само утверждение о том, что все вороны черные, не вытекает из данной конъюнкции, если не принять дополнительную универсальную посылку: « a_1, a_2, \dots включают всех существующих воронов». Кто-то мог бы возразить, что возможность существования каузальных факторов, неизвестных современной физике и не составленных из элементарных частиц, настолько мала, что ею можно пренебречь. Но в этом случае мы покидаем поле логического анализа. Более того, задумавшись на мгновение, мы вспомним, что сегодня нам известно множество каузальных факторов, которые не были бы включены физиком XIX века в опись «содержимого мира». Возможно, атомы и их напоминающие солнечную систему компоненты, электроны и ядра, физик XIX века мог бы предугадать; но что сказать о квантах и гравитационных полях, если окажется, что они существуют? Нет, гипотеза о том, что любая «опись содержимого мира» включает все существующие основные «строительные блоки» каузального процесса, является синтетическим высказыванием и не может считаться истинной с точки зрения чистой логики.

Материализм для современного философа, как я уже отмечал, — это воображаемый противник, которого легко опровергнуть. Современные материалисты (или «теоретики тождества», как они предпочитают себя называть) не утверждают, что *интенционалы* таких слов, как «предпочтение» можно определить в физико-химических терминах, а утверждают лишь, что существует физический референт

этих слов. Их формулировка звучала бы примерно так: предпочтение альтернативы *A* альтернативе *B* *синтетически тождественно* обладанию определенными более или менее устойчивыми особенностями физико-химического строения (например, «предпочтение альтернативы *A* альтернативе *B* есть довольно устойчивое состояние коры головного мозга человека»). Эта позиция сталкивается с той трудностью, что *предпочтение* — это универсалия, а не конкретная сущность (предпочтение альтернативы *A* альтернативе *B* представляет собой *отношение* между организмом и двумя альтернативами, а связка «есть», применяемая к универсалиям, видимо, является связкой «*есть*», используемой при *анализе значения*). Мы говорим, например, что «*растворимость* есть свойство, которым нечто обладает в том и только в том случае, если, будучи помещенным в воду, оно растворяется». Мы *не* говорим, что «растворимость есть определенная физико-химическая структура»; но говорим, что растворимость веществ, которые растворяются, *объясняется* тем, что эти вещества имеют определенную физико-химическую структуру. Аналогичным образом, в случае наших машин мы бы сказали, что предпочитать альтернативу *A* альтернативе *B* значит иметь функцию рационального предпочтения, приписывающую *A* более высокую ценность, чем *B*. Если, помимо этого, мы говорим, что предпочтение альтернативы *A* альтернативе *B* «синтетически тождественно» обладанию определенной физико-химической структурой, скажем, определенной триггерной схемой, то мы берем на себя то, что я назвал бы явным и недостаточно оправданным расширением употребления. Например, если одна и та же машина Тьюринга имеет две довольно разные физические реализации, хотя в обоих случаях одинаковыми оказываются не только функции рационального предпочтения, но и машинные таблицы в целом, то мы должны будем сказать, что «предпочтение альтернативы *A* альтернативе *B* *несколько различается* в случае машины 1 и в случае машины 2». Аналогичным образом, мы должны будем сказать, что и свойство «верить, что» несколько различается в этих двух случаях и т. д. Это можно было бы выразить гораздо яснее, сказав, что в этих двух случаях различна реализация машинной таблицы. Однако здесь есть ряд тонкостей, которые стоит иметь в виду.

Сказанное до этого момента может внушить неверную мысль, что два свойства могут быть тождественными только *аналитически*, но не *синтетически*. Это не так. Допустим, «*a*» — индивидуальная константа, обозначающая конкретный лист бумаги, и на этом листе бумаги я пишу одно слово — «красный». Тогда утверждение «Свойство

красный тождественно свойству, обозначенному единственным словом, написанным на a_1 » является синтетическим ⁷. Однако *только* таким способом можно установить «синтетическую тождественность» свойств; утверждения же «Растворимость есть определенная молекулярная структура», «Боль есть возбуждение C —волокон» не относятся к этому виду, в чем нетрудно убедиться.

До сих пор я предполагал, что, если не считать упомянутый вид синтетических предложений тождества, то критерием *тождества* двух свойств является *синонимичность*, или эквивалентность соответствующих десигнаторов, истолкованная в некотором аналитическом смысле. В статье «Сознание и машины» ⁸ я отмечал, что, видимо, этот критерий некорректен для ряда других абстрактных сущностей, таких, как ситуации, события, и что, возможно, есть основания отказаться от него даже в отношении свойств. В той статье, ссылаясь на связку «*есть*», выражающую *теоретическое тождество* (примерами употребления этой связки служат предложения «вода *есть* H_2O », «свет *есть* электромагнитное излучение»), я предположил, что некоторые свойства можно связывать такого рода связкой «*есть*». Но это не помогло бы теоретику тождества. (Это свидетельствует об изменении моей позиции, изложенной в более ранней работе.) Даже если в некоторых случаях, когда десигнаторы « P » и « Q » не являются синонимами, мы готовы утверждать, что «*быть P есть* (значит) *быть Q* »; для этого требуется, чтобы эти десигнаторы были эквивалентными, а их эквивалентность была *необходимой*, по крайней мере, в смысле *физической необходимости*. Так, если бы оказалось, что *один* конкретный вид физико-химического строения объясняет *все* случаи растворимости, то мы не совершили бы неоправданного расширения обычного употребления этого термина, когда стали бы говорить, что растворимость *есть* обладание этим конкретным видом физико-химического строения. В одной моей более ранней работе обосновывается, что это необязательно было бы «изменением значения». В рассматриваемом случае такое не может иметь места. *Невозможно* открыть законы, устанавливающие, что предпочтение организмом альтернативы A альтернативе B физически необходимо в том и только в том случае, когда этот организм находится в определенном физико-химическом состоянии. Ибо мы уже

⁷ Еще более простой пример: утверждение «синий — это цвет неба» представляет собой синтетическое утверждение о тождестве свойств. Я обязан Нейлу Вильсону из Университета Дюка и этим примером, и разъяснениями по проблеме тождества свойств.

⁸ См. сноску 1 — *Прим. перев.*

знаем, что любой подобный закон был бы ложен. Он был бы ложен, потому что даже в свете нашего современного знания ясно, что любую физически реализуемую машину Тьюринга можно реализовать множеством совершенно разных способов. Поэтому не существует физико-химической структуры, обладание которой было бы необходимым и достаточным условием предпочтения альтернативы *A* альтернативе *B*, даже если мы понимаем «необходимое» не как логически, а как *физически* необходимое. А если бы мы стали говорить о свойствах как «тождественных в некоторых случаях» на том основании, что они имеют одинаковые экстенционалы *в этих случаях*, то это было бы не только изменением значения, но и совершенно произвольным изменением.

До сих пор мы приписывали нашим машинам только такие «многодорожечные» (multi-tracked) предрасположенности, как предпочтения и мнения (beliefs), но речь почти не шла о состояниях с меньшим числом переходов, например, о чувственных состояниях. Конечно, если, снабдив наши машины органами чувств, мы полагаем, что эти органы чувств не являются абсолютно надежными, то, как я отмечал в своей более ранней работе, легко предвидеть, что в «жизни» машины автоматически появится различие между чувственным восприятием (apprehance) и реальностью. Определенные конфигурации в этих машинах можно классифицировать как «зрительные восприятия», «тактильные восприятия» и т. п. Но что можно сказать о таком чувстве как боль?

Адаптируя соответствующим образом анализ, проведенный Стюартом Хэмпширом в книге «Чувство и выражение» (*«Feeling and Expression»*), мы можем внести в нашу модель аналог боли. Идея Хэмпшира состоит в том, что чувства — это состояния, характеризующиеся тем, что они порождают определенные склонности. Например, обычно, хотя и не всегда, причиной боли служит повреждение, нанесенное какой-то части тела, и это порождает склонность отдёргивать эту часть тела, которая представляется поврежденной, и избегать всего, что может вызвать данное болезненное повреждение. Эти склонности в определенном смысле *спонтанны* — это следует подчеркнуть во избежание серьезной критики. Так, если *X* причиняет моей руке боль, склонность отдёргивать руку от *X* появится у меня сразу и без всяких раздумий с моей стороны. Я могу в ответ на вопрос: «Почему вы отдёргнули руку от *X*?», сказать, что «*X* причиняет моей руке боль». В этом случае не последует вопроса: «Почему это причина для отдёргивания руки от *X*?» Тот факт, что боль, которую *X* причиняет моей руке, *ipso facto* служит причиной отдёргива-

ния руки от X , имеет в своей основе и предполагает спонтанную склонность отдёргивать руку от X , находясь в данном состоянии.

Теперь снабдим наши машины «болевыми сигналами», т. е. сигналами, посылаемыми при нанесении повреждения какой-либо части машинного «тела», а также снабдим их «волокнами, проводящими болевые импульсы», и «состояниями боли». Эти «состояния боли» обычно причинно обусловлены повреждениями, нанесенными различным частям машинного тела, и порождают спонтанную склонность избегать всего, что причиняет указанную боль. Думаю, теперь ясно, каким образом можно ввести понятие склонности в нашу модель: склонности естественно трактовать как более или менее кратковременные изменения функции рационального предпочтения машины. На время машина как бы приписывает очень высокую ценность «отдёргиванию руки от того-то». Это *временное* изменение функции рационального предпочтения машины, конечно, не следует путать с долговременным изменением в поведении машины, вызванным уяснением того, что нечто, прежде считавшееся неболезненным, стало причинять боль. Начнем с того, что это последнее изменение может быть встроено в функцию рационального предпочтения машины и для его объяснения не нужно предполагать, что опыт болевых ощущений изменил долговременную функцию рационального предпочтения машины (хотя в динамической модели это может иметь место). В некотором смысле это вносит усложнение в модель Хэмпшира⁹: для описания состояний боли используются как кратковременные спонтанные склонности, порождаемые этими состояниями, так и негативное значение, приписываемое основной функцией рационального предпочтения машины вещам, которые, как известно машине из опыта, подвергают ее в эти состояния.

Выдвинутые выше соображения против отождествления предпочтения с конкретным физико-химическим строением в равной мере направлены и против отождествления боли с конкретным физико-химическим строением. Предположим, что волокна машины, проводящие болевые импульсы, сделаны из меди, и это единственные медные волокна в машине. Было бы абсурдно утверждать, что «Боль есть возбуждение медных волокон». Утверждая это, мы были бы вынуждены признать, что если машина 1 имеет медные волокна, а машина 2 — платиновые, то боль — это нечто разное в случае машины 1 и в случае машины 2. К тому же, мы выразимся более ясно, если скажем, что «боль *есть* состояние машины, обычно возникаю-

⁹ Хотя здесь не обсуждаются другие аспекты модели Хэмпшира: роль *необусловленных* реакций, «подавление» склонностей и роль имитации.

щее при нанесении повреждений телу машины и характеризуемое тем, что оно порождает «склонности к...» и т. п., и поэтому следует отказаться от формулировки: «Боль синтетически тождественна возбуждению медных волокон» в пользу более ясной формулировки «Машина физически реализована таким образом, что «болевые» импульсы распространяются по медным волокнам».

3. ЛОГИЧЕСКИЙ БИХЕВИОРИЗМ

Как мы видели, утверждения о предпочтениях наших машин логически не эквивалентны утверждениям об их физико-химическом строении. Быть может, они логически эквивалентны утверждениям о реальном и потенциальном поведении этих машин? Для ответа на этот вопрос имеет смысл расширить рамки обсуждаемой темы, включив в нее не только утверждения о предпочтениях наших машин, но и утверждения об их «знании», «мнении» и «чувственном знании» (*sensory awareness*). Расширив таким образом рамки нашего обсуждения, мы сразу поймем, что ответ будет негативным. Рассмотрим две машины T_1 и T_2 , имеющие следующее различие: у T_1 удалены «волокна, проводящие болевые импульсы», поэтому T_1 неспособна «ощущать боль». У T_2 эти волокна не удалены, но она имеет необычную функцию рационального предпочтения. Эта функция такова, что если T_2 считает, что имело место определенное событие или что истинно определенное высказывание, то T_2 припишет *бесконечно большое значение* сокрытию того факта, что у нее не удалены волокна, проводящие болевые импульсы. Другими словами, будучи спрошенной, T_2 будет утверждать, что ее волокна были удалены и она неспособна ощущать боль, и будет подавлять в себе склонность обнаружить в поведении ощущаемую боль. Если T_2 не считает, что какое-то важное событие имело место или какое-то важное высказывание истинно, то у T_2 не будет причин для сокрытия того, что она способна «ощущать боль», и, поэтому, она будет вести себя совсем иначе, чем T_1 . В этом случае, наблюдая поведение машины, мы сможем установить, является ли она физической реализацией T_2 или T_1 .

Однако, если и T_1 , и T_2 известно о том, что имело место какое-то важное событие, или истинно какое-то важное высказывание, то их нельзя будет отличить по их поведению. Стало быть, если и машина T_1 и машина T_2 считают, что имело место какое-то важное событие или истинно какое-то важное высказывание, то гипотеза о том, что какая-то конкретная машина — это реализация T_1 , а гипотеза о том, что эта машина относится к типу T_2 , будут совпадать в своих предсказаниях относительно реального и потенциального по-

ведения этой машины. Короче говоря, определенные комбинации довольно разных мнений и функций рационального предпочтения будут иметь следствием совершенно одинаковое поведение.

В статье «Мозг и поведение»¹⁰ я доказывал, что то же самое истинно и в отношении людей. Так, два человека могут вести себя совершенно одинаково при всех возможных обстоятельствах, один — в силу обычных причин, а другой — в силу довольно необычной совокупности причин. Если допустить разные вычислительные способности или разный интеллект у машин, то всё станет еще более очевидным. Рассмотрим случай, когда нам нужно отличить машину со стандартной функцией рационального предпочтения и достаточно низким интеллектом от машины с очень высоким интеллектом и нестандартной функцией рационального предпочтения, приписывающей бесконечно большое значение сокрытию высокого интеллекта. Ясно, что различие между ними не является абсолютно невыявляемым. Если бы нам разрешили разобрать эти машины и установить, что происходит у них внутри, то мы смогли бы сказать относительно какой-то конкретной машины, принадлежит ли она к первому типу или ко второму, но нетрудно показать, что без изучения внутреннего строения рассматриваемых машин невозможно отличить их друг от друга. Стало быть, довольно разные комбинации вычислительных способностей, мнений и функций рационального предпочтения могут иметь следствием совершенно одинаковое поведение — не только совершаемое в настоящий момент, но и потенциальное поведение при всех возможных обстоятельствах.

Пусть T_1 — машина с низким интеллектом, а T_2 — машина с высоким интеллектом, и T_2 имитирует поведение T_1 . Кто-то мог бы спросить, в чем же проявляется более высокий интеллект машины T_2 ? Он проявляется в двух вещах. Прежде всего, T_2 может печатать на своей ленте много такого, что не содержит сигналов-команд и что, следовательно, просто составляет ее внутренний монолог. T_2 умеет решать математические задачи, анализировать психологию людей, с которыми она общается, записывать едкие замечания по поводу человеческих нравов и институтов, и т. п. T_2 вовсе не нужно содержать какую-либо подсистему состояний, хоть сколько-нибудь совпадающих с состояниями или вычислениями T_1 . Возможно, T_2 достаточно умна, чтобы определить, как поступила бы T_1 в той или иной конкретной ситуации, и ей не нужно воспроизводить ход рас-

¹⁰ Putnam H. Brains and Behavior // Mind, Language and Reality, Philosophical Papers. Vol. 2. Cambridge: Cambridge University Press, 1975, p. 325—341.

суждений, приведший T_1 к решению поступить таким-то образом. Мы имели бы аналогичный случай, если бы поведение какого-то человека ничем не отличалось от поведения других людей, но втайне от всех этот человек имел бы богатую и необычную внутреннюю жизнь.

Нам укажут, что все рассмотренные нами машины имеют, в некотором смысле, *патологические функции* рационального предпочтения, т. е. функции, которые приписывают каким-то вещам бесконечно большое значение. Приписывать какой-либо вещи бесконечно большое значение значит предпочитать эту вещь всем другим возможным альтернативам, к чему бы это не привело. Назовем функцию рационального предпочтения *непатологической*, если она никому *не* приписывает бесконечно большого значения, за исключением, пожалуй, выживания самой машины. Пусть T — теория, согласно которой все реально существующие интеллектуальные системы имеют непатологические функции рационального предпочтения. Тогда можно показать, что в теории T утверждение о том, что машина с такими-то зафиксированными вычислительными способностями имеет такую-то определенную функцию рационального предпочтения, эквивалентно утверждению, что эта машина демонстрирует определенное поведение. По сути, утверждение о том, что машина имеет какую-то определенную функцию рационального предпочтения, эквивалентно в теории T утверждению о том, что при всех обстоятельствах она ведет себя точно так же, как вела бы себя машина, имеющая эту функцию рационального предпочтения. Тем не менее, это не доказывает истинности логического бихевиоризма, хотя он почти «попадает в цель». Применительно к нашим машинам логический бихевиоризм формулировался бы так: утверждение о том, что машина имеет какую-то определенную функцию рационального предпочтения, логически эквивалентно утверждению о ее реальном и потенциальном поведении. Это неверно. Верно же то, что имеется теория T , истинная с высокой степенью вероятности (или с высокой степенью вероятности истинен ее аналог, сформулированный для организмов), такая, что в теории T каждое утверждение, имеющее форму « T предпочитает альтернативу A альтернативе B », эквивалентно утверждению о реальном и потенциальном поведении T . Но эквивалентность как чисто логический вопрос и эквивалентность в рамках синтетической теории — это две совершенно разные вещи.

В некотором смысле логический бихевиоризм находится в таком же положении, что и материализм. Как мы видели в случае материализма, хотя утверждение о том, что машина имеет определенную

машинную таблицу, логически не эквивалентно утверждению о том, что ей присуще определенное физико-химическое строение, однако, первое утверждение вытекает из второго в рамках некоторой синтетической теории, а именно, теории, которая включает в себя законы природы и тезис о полноте, т. е. тезис о том, что не существует никаких иных каузальных факторов, помимо элементарных частиц и соединений элементарных частиц, и они обладают только той степенью свободы, которую приписывает им физическая теория. Действительно, нетрудно убедиться в том, что имеется класс физических структур C такой, что утверждение о том, что машина имеет определенную машинную таблицу, эквивалентно в рамках вышеупомянутой синтетической теории утверждению о том, что физическая структура этой машины принадлежит к классу C . Поскольку утверждения о том, что машина предпочитает альтернативу A альтернативе B , придерживается определенного мнения или «ощущает боль» и т. п., истинны только в том случае, если истинна соответствующая конъюнкция, состоящая из двух утверждений, одно из которых говорит, что машина имеет определенную машинную таблицу, а второе описывает полную конфигурацию машины в данный момент времени, и поскольку подобная конъюнкция может быть истинной в рамках вышеупомянутой синтетической теории только в том случае, если физическая структура машины принадлежит к очень широкому классу физических структур C^* , то любое из перечисленных утверждений эквивалентно в рамках вышеупомянутой синтетической теории утверждению, что физическая структура машины принадлежит к этому классу C^* .

Аналогичным образом, если принять синтетическую теорию, о которой шла речь в связи с логическим бихевиоризмом, и которая утверждает, что ни одна машина не обладает патологической функцией рационального предпочтения, то любое утверждение о «ментальной жизни» любой из наших машин будет эквивалентно соответствующему утверждению о реальном и потенциальном поведении этой машины.

Применительно к «агенту» нашего гипотетического «сообщества» ситуацию можно описать так: располагая достаточной информацией о реальном и потенциальном поведении агента, мы можем с относительной достоверностью сделать вывод о том, что для агента предпочтительнее A , чем B , или, с другой стороны, располагая достаточной информацией о физико-химической структуре агента (и достаточными знаниями о законах природы), мы также можем с относительной определенностью сделать вывод о том, что для агента

предпочтительнее *A*, чем *B*. Однако эти два вывода не служат подтверждением ни логическому бихевиоризму, ни материализму, соответственно. Оба вывода представляют собой синтетические выводы, полученные в рамках синтетических теорий.

Однако правомерен вопрос: Как вообще мы можем знать, что предположение о несуществовании патологических функций рационального предпочтения и предположение о полноте физической теории правильны? Думаю, что в обоих случаях ответ практически одинаков. Каждое из этих предположений оправданно, поскольку нет серьезных оснований считать его ложным. Если так, то выводы о ментальной жизни любой реальной эмпирической системы могут быть совершенно оправданными; но, они никоим образом не будут аналитическими выводами, если в посылах содержится информация только о реальном и потенциальном поведении системы и о ее физико-химическом строении. Такие выводы можно всегда «аннулировать», поскольку всегда есть нестандартные обстоятельства, при которых посылки могут выполняться, а заключение — быть ложным.

Просмотрев все, что я здесь написал, я, признаюсь, испытываю некоторое разочарование. Все эти вещи представляются мне слишком очевидными и тривиальными, чтобы о них говорить, хотя некоторые философы и не согласились бы с ними. Вместе с тем, думаю, что высказанные здесь соображения при всей их кажущейся очевидности позволяют увидеть в характере наших менталистских понятий нечто такое, на что обычно совсем не обращают внимания. Представляется, что если мы достоверно знаем, что человек придерживается каких-то определенных мнений, предпочтений и т. д., то мы знаем что-то о его функциональной организации. В случае машин Тьюринга функциональная организация задается машинной таблицей. Описание функциональной организации человека может быть совершенно иным и куда более сложным. Но важно то, что описания функциональной организации системы по своему типу логически отличаются как от описаний ее физико-химического строения, так и от описаний ее реального и потенциального поведения. Думаю, что исследования в философии сознания потому так часто удивляют своей неудовлетворительностью, что в них не учитывается понятие функциональной организации или же оно смешивается с понятиями совершенно иного рода.

ФИЛОСОФИЯ И НАША МЕНТАЛЬНАЯ ЖИЗНЬ

Вопрос, который давно волнует и простых людей, и философов, хотя современный аналитический стиль философствования скрывает это обстоятельство, звучит так: состоим ли мы из материи или из духовной субстанции (*soul-stuff*)? Или, говоря прямо, являемся ли мы просто материальными существами или «чем-то бóльшим»? В настоящей статье я попытаюсь как можно убедительнее показать, что в основе этого вопроса лежит ложная посылка. Однако моя цель — не столько устранить этот вопрос, сколько высказаться по реальной проблеме, которая за ним стоит. Думаю, что этой проблемой является автономия нашей ментальной жизни.

Людей беспокоит возможность разоблачения, когда окажется, что наше поведение объясняется чем-то механическим. Разумеется, я понимаю механическое не в старом смысле зубцов и шестеренок, а в новом смысле электричества и магнетизма, квантовой химии и т. п. В настоящей статье я, в частности, попытаюсь доказать, что такое разоблачение невозможно в принципе. Ментальность представляет собой реальную и автономную характеристику нашего мира.

Однако еще более важное значение, по крайней мере на мой взгляд, имеет то, что в целом этот вопрос не имеет никакого отношения к нашей субстанции. Каким бы странным это ни казалось и с точки зрения здравого смысла, и с точки зрения утонченной интуиции, но вопрос об автономии нашей ментальной жизни никак не связан с весьма распространенным и старым вопросом о материи и духовной субстанции. Мы могли бы быть сделаны и из швейцарского сыра, но это не имело бы никакого значения.

Неспособность понять это, упрямое и настойчивое стремление ставить вопрос именно так: *материя или душа?* — крайне мешают прогрессу в решении этой проблемы. Но, как только мы осознаем, что проблема заключается не в нашей субстанции, то сможем значительно продвинуться вперед.

На мой взгляд, основным понятием, которое поможет раскрыть тайны в философии сознания, является понятие *функционального изоморфизма*. Две системы функционально изоморфны, если между

* Эта статья была представлена как доклад на Фёрстеровском симпозиуме по теме «Компьютеры и мышление» в Калифорнийском университете (Беркли) в октябре 1973 г. Я благодарен Алану Гарфинкелю за полезные замечания, сделанные по поводу более ранней версии этой статьи.

состояниями одной и состояниями другой существует соответствие, предполагающее сохранение функциональных отношений. Для начала возьмем пример с компьютерами: пусть функциональные отношения — это просто отношения «следования за», например, за состоянием A всегда следует состояние B ; чтобы F было функциональным изоморфизмом, должно выполняться следующее: состояние B следует в системе 1 за состоянием A , тогда и только тогда, когда в системе 2 за состоянием $F(A)$ следует состояние $F(B)$. Если функциональные отношения касаются, скажем, данных, печатаемых на ленте, то когда печатается на ленте знак π , система 1 переходит в состояние A , и если существует F -функциональный изоморфизм между системой 1 и системой 2, то эти отношения должны сохраняться, т. е. когда печатается на ленте знак π , система 2 переходит в состояние $F(A)$. В более общем виде, если T — правильная теория, описывающая функционирование системы 1 на функциональном или психологическом уровне, то изоморфизм между системой 1 и системой 2 означает, что каждое свойство и отношение, определяемое в системе 2, должно отображаться таким образом, что T принимает истинное значение, когда все ссылки на систему 1 заменяются ссылками на систему 2, а все символы свойств и отношений в T получают переинтерпретацию в соответствии с указанным отображением.

С понятием функционального изоморфизма связана та трудность, что оно предполагает понятие функционального или психологического описания. Именно по этой причине в других своих статьях я использовал машины Тьюринга для разъяснения понятия функционального изоморфизма. Таким образом, я вынужден был отстаивать тезис, что мы представляем собой машины Тьюринга. Для функционального описания машины Тьюринга имеется нормальная форма, так называемая машинная таблица — стандартный вид программы. Однако мне не кажется фатальной небрежностью, хотя в этом есть доля небрежности, — применять понятие функционального изоморфизма к системам, в отношении которых мы в настоящий момент не можем сказать, как выглядела бы нормальная форма для их описания, — к системам типа нас самих. Дело в том, что хотя мы и не представляем, как должна выглядеть исчерпывающая психологическая теория, я убежден, что мы знаем достаточно (и в данном случае нам служат подспорьем аналогии с компьютерами, экономическими системами, играми и т. п.), чтобы указать четкие различия между любой возможной психологической теорией человека, функциональными описаниями компьютера и экономической систе-

мы, с одной стороны, и их физическим или химическим описанием, с другой. И действительно, Деннет и Фодор в своих последних книгах намного продвинулись в этом направлении.

Все это возвращает меня к вопросу о *меди, сыре и душе*. Поскольку у нас есть базовое понятие функционального изоморфизма, один момент мы можем отметить сразу же: две системы могут иметь достаточно разное строение и быть функционально изоморфными. Например, компьютер, изготовленный из электротехнических деталей, может быть изоморфным компьютеру, изготовленному из зубцов и шестеренок. Иными словами, каждому состоянию в первом компьютере соответствует определенное состояние во втором, и, как мы уже видели, отношения «следования за» у них одинаковы: если в случае электронного компьютера за состоянием *S* следует состояние *B*, то в случае компьютера, изготовленного из зубцов и шестеренок, за состоянием *A* будет следовать состояние *B*; и не имеет никакого значения то, что *физическая реализация* этих состояний совершенно различна. Следовательно, компьютер, изготовленный из электротехнических деталей, может быть изоморфным компьютеру, изготовленному из зубцов и шестеренок, или клерку, пользующемуся бумагой и ручкой. Производя вычисления, компьютер, изготовленный из какого-то сорта проволоки, скажем, медной проволоки, и какого-то сорта реле и т. д., будет находиться в ином физическом и химическом состоянии, чем компьютер, изготовленный из другого сорта проволоки и других реле. Однако их функциональное описание может быть одинаковым.

Эту мысль можно развить дальше. Предположим, что один из тезисов материализма (назовем его «первым тезисом») верен, и мы, в целом, — просто материальные системы, подчиняющиеся физическим законам. Тогда второй тезис классического материализма не может быть верным, то есть наши ментальные состояния, например, *мысли о будущих летних каникулах*, не могут быть *тождественными* какому-либо физическому или химическому состоянию. Из того, что нам уже известно о компьютерах и т. п., ясно, что какой бы ни была программа мозга, у нас есть физическая возможность, пусть и не реализуемая с необходимостью, создать то, что имеет ту же самую программу, но обладает иным физическим и химическим строением. Но тогда было бы абсурдно отождествлять рассматриваемое состояние с его физической или химической реализацией, учитывая, что в определенном смысле эта реализация довольно случайна, во всяком случае, с точки зрения психологии (науки, имею-

щей к этому прямое отношение) ¹. Это как если бы мы встретили марсиан и, обнаружив, что они во всех функциональных аспектах изоморфны нам, отказывались бы признать, что они могут ощущать боль, потому что у них другие С-волокна.

Теперь представим себе две возможные вселенные, например, два «параллельных мира» в научно-фантастическом смысле, в одном из которых люди в старом добром смысле наделены душой, локализованной в шишковидной железе, а в другом — люди имеют сложно устроенный мозг. Предположим, что души из первого мира функционально изоморфны мозгам из второго мира. Нужно ли придавать этому различию большее значение, чем различию между медной и немедной проволокой в случае компьютера? Важно ли то, что люди, наделенные душой, имеют, так сказать, нематериальный мозг, а люди, наделенные мозгом, имеют материальную душу? Значение имеет только общая структура, т. е. теория *T*, которая, увы, нам совершенно неизвестна, а не *техническое обеспечение* (hardware), будь оно даже бесплотным.

Против сказанного мною можно выдвинуть множество возражений. Я попытаюсь ответить на некоторые из них.

Например, можно было бы сказать, что если души людей, наделенных душой, изоморфны мозгам людей, наделенных мозгом, то эти души должны быть подобны автоматам, а это не тот сорт души, который нас интересует. «Ваш аргумент показывает лишь, что не нужно проводить различие между мозгом и автоматоподобной душой». Но что именно означает это возражение?

Думаю, есть два способа его истолкования. Это возражение может означать, что понятие функциональной организации или функционального изоморфизма имеет смысл только для автоматов. Но это совершенно неверно. Какими бы неадекватными ни были наши сегодняшние понятия, мы, как подчеркнул Джерри Фодор, по крайней мере знаем, что понятие функциональной организации применимо ко всему, к чему применимо понятие психологической теории. Разъясняя наиболее общее понятие функционального изоморфизма, я указывал, что две системы функционально изоморфны, ес-

¹ Даже если бы было физически невозможно воссоздать человеческую психологию в существе, состоящем не из обычной протоплазмы, ДНК и т. д., то все равно было бы неверно говорить, что психологические состояния тождественны со своими физическими реализациями. Поскольку, как мы увидим ниже, подобное отождествление не имеет *объяснительной* ценности в психологии. По этому вопросу см. работу: Fodor J. Psychological Explanation. N. Y., 1968.

ли имеется изоморфизм, благодаря которому они выступают моделями одной и той же психологической теории. (Это более сильное утверждение, чем просто сказать, что они обе являются моделями одной и той же психологической теории, ибо они являются изоморфными реализациями одной и той же абстрактной структуры.) Утверждать, что старая добрая душа не попадает в область определения понятия функциональной организации или понятия функционального изоморфизма, значит вставить на ту точку зрения, что для души не может существовать теории, что бы под ней мы ни понимали. Думаю, это откровенный обскурантизм. Поэтому я буду исходить из того, что понятие сознания или души и т. п. не предполагает, что они непостижимы или для них нельзя построить теорию.

Во-вторых, кто-то мог бы выдвинуть более серьезное возражение, сказав, что даже если и существует теория души или сознания, то душа, по крайней мере в ее широком, привычном смысле, обладает такими способностями, какие не могла бы иметь механическая система. Я рассмотрю это утверждение в заключительной части настоящей статьи.

Если для кого-то понятие души предполагает, что душа может совершать вещи, нарушающие законы физики, то признаю, это ставит меня в тупик. Душа не может быть изоморфной мозгу, если она способна к ясновидению, которое нельзя объяснить с помощью законов физики. С другой стороны, если кого-то интересуют более простые формы магии, такие как телепатия, то мне думается, что в принципе не может быть причин, запрещающих создать устройство, передающее мысленные сообщения от одного мозга к другому. Что касается перевоплощения, то если мы, как я утверждаю, представляем собой определенный вид функциональные структуры (как таковая моя идентичность — это моя функциональная структура), то думаю, что в принципе нет причин, почему эта структура не могла бы быть воспроизведена через тысячу, миллион или миллиард лет. О Воскрешении: христиане, как вы знаете, верят в воскрешение во плоти, которое совершенно не предполагает нематериального носителя. Поэтому при рассмотрении даже этих вопросов (они не составляют предмет исследования в настоящей статье, хотя я заинтересован в диалоге с теми, для кого они представляют интерес), нам не нужен нематериальный мозг или духовная субстанция.

Итак, если я прав, и вопрос о материи и духовной субстанции действительно не имеет философской или религиозной значимости, то почему ему уделяется столько внимания, почему он так волнует людей? Видимо, все дело в том, что и сторонники Дидро, и сторон-

ники Декарта согласились бы с тем, что если мы состоим из материи, то существует физическое объяснение нашему поведению, но у одних это вызвало бы разочарование, а у других — восторг. Думаю, традиционный дуалист воскликнул бы: *«Разве не ужасно, если бы оказалось, что мы все состоим из одной материи, ведь тогда есть физическое объяснение всему, что мы делаем»*. А традиционный материалист сказал бы: *«Если мы лишь материя, то существует физическое объяснение для всего, что мы делаем. Разве это не восхитительно!»* (Это подобно различию между оптимистом и пессимистом: оптимист — это человек, который говорит: «Это лучший из миров»; а пессимист ему отвечает: «Увы, вы правы».)²

Думаю, что и те, и другие неправы. И Дидро, и Декарт ошибались, думая, что если мы состоим из материи и если наши души материальны, то существует физическое объяснение нашего поведения.

Позвольте мне пояснить свою мысль очень простой аналогией. Допустим, что у нас есть крайне простая физическая система — доска с двумя отверстиями в виде круга диаметром в один дюйм и в виде квадрата со стороной в один дюйм и кольцо кубической формы со стороной, равной $15/16$ дюйма. Нам нужно объяснить очень простой факт: *почему кольцо проходит через квадратное отверстие, но не проходит через круглое*.

Мы могли бы попытаться объяснить это следующим образом: кольцо представляет собой облако или, лучше сказать, жесткую решетку атомов. Можно было бы даже попытаться дать описание этой решетки, вычислить ее потенциальную электрическую энергию, задаться вопросом о том, почему она не разрушается, привлечь квантовую механику для объяснения ее устойчивости и т. п. Доска также представляет собой решетку атомов. Назовем кольцо — «системой А», а отверстия — «зоной 1» и «зоной 2». Мы могли бы вычислить все возможные траектории движения системы А (кстати, в ходе этих вычислений возникают очень серьезные проблемы относительно их эффективности, осуществимости и т. д., но в данном случае мы это опустим). Вероятно, мы могли бы из одних только законов механики элементарных частиц или квантовой электродинамики вывести, что система А никогда не пройдет через зону 1, но существует по крайней мере одна траектория, по которой она может пройти через зону 2. Но будет ли это объяснением того, почему ко-

² Шутка заимствована у Джозефа Вайценбаума.

лышек проходит через квадратное отверстие и не проходит через круглое?

Очень часто мы слышим, что если что-то состоит из материи, то его поведение должно иметь физическое объяснение. Это доказываете тем, что если что-то состоит из материи (и тут мы принимаем множество допущений), то его поведение должно быть дедуцируемо из его материальной структуры. *Что заставляет нас считать эту дедукцию объяснением?*

С другой стороны, если вы не «помешаны» на той идее, что объяснение должно касаться уровня элементарных компонентов, и признаете такую возможность, что для объяснения *имеют значение не элементарные компоненты, а только структуры высокого уровня*, то нашему примеру есть очень простое объяснение: доска и колышек — это жесткие предметы, и, согласно геометрии, круглое отверстие меньше, а квадратное отверстие больше поперечного сечения колышка. Колышек проходит через отверстие, достаточно большое, чтобы вместить его поперечное сечение, и не проходит через отверстие, слишком маленькое, чтобы вместить его поперечное сечение. Это объяснение правильно, независимо от того, состоит ли колышек из молекул, представляет ли собой сплошную жесткую субстанцию или что-либо еще. (Для детализации этого объяснения можно было бы отметить тот геометрический факт, что квадрат со стороной в один дюйм больше круга диаметром в один дюйм.)

Итак, можно сказать, что это объяснение выявляет определенные *значимые (relevant) структурные характеристики рассматриваемой ситуации*, а именно — геометрические характеристики. *Имеет значение* то, что квадрат со стороной в один дюйм больше круга диаметром в один дюйм. И отношение между размером и формой колышка, с одной стороны, и размером и формой отверстия, с другой, *имеет значение*. Имеет значение и то, что колышек и доска сохраняют *жесткость* при перемещении. Ничто другое значения не имеет. Это же объяснение будет правильным в любом другом мире (какова бы ни была его микроструктура), где имеют место *эти высокоуровневые структурные характеристики*. И в этом смысле *данное объяснение автономно*.

По мнению некоторых людей, я неправ, утверждая, что дедукция из микроструктуры явления не является его объяснением. Думаю, если принять во внимание, в каких *целях мы используем понятие объяснения*, она не является объяснением. Дедукция, если угодно, *есть* объяснение, но неудовлетворительное объяснение. И

непонятно, зачем оно нужно, если можно дать удовлетворительное объяснение?

Удовлетворительность объяснения — это не субъективная оценка. Даже если не соглашаться с позитивистами, навязывающими нам понятие объяснения как дедукции из законов, поиск законов все равно представляет собой одну из целей науки. Объяснение, если оно выявляет соответствующие законы, не только в субъективном, но и в *методологическом отношении* лучше, ибо оно способствует достижению целей научного исследования. Объяснение лучше, если оно имеет более общий характер.

Возьмем только эти две характеристики, хотя можно привести и много других, и сравним два объяснения явления: с учетом более высокого уровня его структуры и с учетом его атомарной структуры. Объяснение на более высоком уровне выявляет соответствующие геометрические соотношения. Объяснение на более низком уровне скрывает эти законы. Отметим также, что объяснение на более высоком уровне применимо к гораздо более интересному классу систем (конечно, это зависит от того, что нас интересует).

Дело в том, что мы гораздо более заинтересованы в обобщениях, применимых к другим жестким структурам, демонстрирующим разнообразные геометрические соотношения, чем в обобщениях, применимых к *следующему колышку, обладающему точно такой же молекулярной структурой*, по той простой причине, что не существует другого колышка, обладающего точно такой же молекулярной структурой. Поэтому для имеющих практическое значение научных дисциплин и для практических способов решения научных проблем объяснение с учетом более высокого уровня структуры явления носит гораздо более общий характер, а, стало быть, и *объясняет*.

Только с учетом более *высокого* уровня структуры мы смогли дедуцировать законообразное утверждение о том, что колышек проходит через отверстие, которое больше его поперечного сечения. Когда мы пытаемся дедуцировать возможные траекторию движения «системы А» из утверждений об индивидуальных атомах, мы используем совершенно случайные посылки: этот атом — здесь, тот атом кислорода — там и т. д. и это одна из причин, почему ошибочно сводить объяснения в науках типа экономики на уровень элементарных частиц, из которых составлены участники экономической игры. На самом деле, их действия — покупка, продажа, установление средней цены — не дедуцируемы из одних только уравнений

движения. Иначе установление средней цены было бы вызвано *физической*, а не *экономической необходимостью*. Люди участвуют в этой игре, потому что они — особые системы, имеющие особые граничные условия, совершенно случайные с точки зрения физики. Это означает, что выведение законов экономики из *одних только* законов физики невозможно *в принципе*. Выведение законов экономики из законов физики и *случайных по характеру утверждений о том, каким частицам где и когда случилось быть*, в принципе, посильно лапласовскому сверхразуму, но зачем оно нужно? Несколько глав, например из фон Неймана, скажут нам гораздо больше об экономических закономерностях, чем любая подобная дедукция.

Из сказанного я делаю следующий вывод: мы, действительно, обладаем тем видом автономии, к которому стремимся в сфере ментального. Какой бы ни была наша ментальная жизнь, видимо, нет серьезных оснований считать, что ее можно *объяснить* с помощью нашего физического и химического строения. Нас здесь интересует не вопрос о том, можно ли было бы, учитывая, что мы состоим из таких-то частиц, предсказать нашу ментальную жизнь, ибо такое предсказание ничего бы не *объяснило*, каким бы великим научным достижением оно ни было. Нас интересует вопрос о том, можно ли с учетом этого автономного уровня и на том основании, что мы имеем такую-то структуру или такую-то программу, вывести, что мы сможем выучить то-то, будем склонны любить то-то и т. д. Все эти проблемы ментальной жизни связаны с описанием автономного уровня ментальной жизни, и именно этот уровень предстоит открыть.

В более ранних статьях я обосновывал гипотезу о том, что: (1) в целом человек — это машина Тьюринга и (2) психологические состояния человека — это состояния машины Тьюринга или дизъюнкции состояний машины Тьюринга. В настоящей статье я хотел бы показать, что эта точка зрения была совершенно ложной и я был слишком увлечен редукционистским подходом.

Начнем с технических трудностей. *Состояние* машины Тьюринга описывается таким образом, что в любой данный момент времени она может находиться только в одном состоянии. Более того, в модели машины Тьюринга запоминание и обучение представляют собой не приобретение новых состояний, а усвоение новой информации, печатаемой на машинной ленте. Таким образом, если люди вообще могут находиться в состояниях, сходных с состояниями машины Тьюринга, то этими состояниями должны быть (1) состояния, в которых человек может находиться в любой момент и на которые не

вливают обучение и память; и (2) *совокупные* мгновенные состояния человека, которые вместе с обучением и памятью определяют, каким будет следующее состояние, и вместе с тем полностью описывают текущее состояние человека («полностью» с точки зрения психологической теории).

Эти условия означают, что *ни одно* психологическое состояние, в его обычном истолковании, не может быть состоянием машины Тьюринга. Возьмем конкретный вид боли в качестве «психологического состояния». Если я *являюсь* машиной Тьюринга, то мое настоящее «состояние» определяется не только тем, ощущаю ли я эту конкретную боль, но и тем, собираюсь ли я произнести слово «три», слышу ли я пронзительный вой и т. д. Поэтому рассматриваемое психологическое состояние (боль) и мое состояние, в смысле *машинного состояния*, — это не одно и то же, хотя (все же) возможно, что мое машинное состояние *определяет* мое психологическое состояние. Более того, *ни одна* психологическая теория не стала бы утверждать, что свойства «ощущать конкретную боль», «собираться произнести слово “три”», «слышать пронзительный вой» и т. д. относятся к *одному* психологическому состоянию, хотя вполне может быть машинное состояние, которое характеризуется тем, что я находился бы в нем, если бы одновременно я ощущал указанную боль, собирался произнести слово «три», слышал пронзительный вой и т. д. Поэтому, даже если я и машина Тьюринга, мои машинные состояния и мои психологические состояния — это *не* одно и то же. Описать меня *как* машину Тьюринга (с помощью машинной таблицы) и описать меня *как* человека (с помощью психологической теории) — значит сделать описания на двух совершенно разных уровнях организации.

Пока нам ничто не запрещает считать психологическое состояние длинной (практически почти бесконечной) дизъюнкцией машинных состояний, несмотря на то, что каждое *отдельное* машинное состояние не является психологическим состоянием. Но мы убедимся в том, что это очень маловероятно, если перейдем от рассмотрения таких состояний, как «боль» (которые являются почти *биологическими*), к рассмотрению состояний типа «ревность», «любовь» или «соперничество». Состояние ревности, безусловно, не является мгновенным, и оно зависит от имеющейся у человека информации, от известных ему фактов и от присущих ему привычек, тогда как состояния машин Тьюринга мгновенны, на них не влияют обучение и память. Так, благодаря обучению и памяти машина Тьюринга может перейти в какое-то состояние, но специфика самого этого со-

стояния не зависит от обучения и памяти, однако если я нахожусь в состоянии, которое можно охарактеризовать так: «Я ревную X к Y », то само это состояние предполагает мое знание о том, что X и Y — люди, что между людьми бывают разного рода отношения и т. д. Поэтому ревность не может быть ни машинным состоянием, ни дизъюнкцией машинных состояний.

Можно было бы попытаться изменить нашу теорию так: ревновать = или находиться в состоянии A и иметь ленту c_1 или находиться в состоянии A и иметь ленту c_2 , или.. находиться в состоянии B и иметь ленту d_1 , или находиться в состоянии B и иметь ленту d_2 ... находиться в состоянии Z и иметь ленту y_1 , или находиться в состоянии Z и иметь ленту y_n , т. е. попытаться определить психологическое состояние как дизъюнкцию, отдельными членами которой являются не состояния машин Тьюринга, как раньше, а конъюнкции машинного состояния и ленты (т. е. полного описания содержимого банка данных). Подобное описание было бы практически бесконечным, но, главное, теория утратила бы смысл, ибо наша первоначальная цель состояла в том, чтобы использовать машинную таблицу в качестве модели психологической теории, а теперь стало ясно, что описание машинной таблицы, хотя и отличается от описания на уровне элементарных частиц, но столь же далеко от описания *с помощью* психологической теории, как и физико-химическое описание.

В чем же состоит значение машин для философии сознания? Думаю, значение машин имеет как положительную, так и отрицательную стороны. Положительная сторона заключается в том, что понятие функциональной организации впервые было введено благодаря машинам, в частности вычислительным машинам. Машины заставили нас осознать важность различия между абстрактной структурой и ее конкретной реализацией. Конечно, это различие было известно и до машин. Но в случае вычислительных машин мы напрямую столкнулись с тем фактом, что фактически одну и ту же структуру можно реализовать невероятно разнообразными способами, что физико-химические свойства не являются важнейшими для машин. Машины заставили нас понять исключительную важность идеи функциональной организации. Вместе с тем, отрицательное значение машин заключается в том, что они толкают нас к упрощениям. Функциональная организация была изучена на примере системы с очень ограниченной и специфической функциональной организацией. Поэтому очень соблазнительно предположить, что и мы должны об-

ладать такой же ограниченной и специфической функциональной организацией.

Сейчас я хотел бы рассмотреть пример, который может оказаться полезным, хотя кому-то он покажется далеким от обсуждаемых здесь проблем. Этот пример не из философии сознания. Рассмотрим такой факт. Земля вращается вокруг Солнца не по кругу, как считали раньше, а по эллипсу, и Солнце находится не в центре, а в одном из фокусов этого эллипса. Однако, как это ни удивительно, сохраняет истинность утверждение, которое было бы истинно, если бы орбита была кругом, а Солнце находилось в центре круга. Это утверждение таково: радиус-вектор, проведенный от Солнца до Земли, проходит равные площади за равные промежутки времени. Совершенно очевидно, что если бы орбитой Земли был круг, то Земля двигалась бы с постоянной скоростью. Но Земля движется не по кругу. И скорость ее движения также непостоянна: когда Земля находится дальше от Солнца, она движется медленнее, а когда она находится ближе к Солнцу, то — быстрее. Земля то ускоряет, то замедляет свое движение. И тем не менее, радиус-вектор Земли проходит равные площади за равные промежутки времени³. Ньютон вывел этот закон в своих «Началах», и это выведение показывает, что для этого закона важно то, что на Землю действует сила, направленная к Солнцу. Для выведения этого закона необходим исключительно этот факт. В математическом плане он равносителен этому закону⁴. Все прекрасно, если притяжение между любыми двумя телами определяется согласно обратно квадратичному закону, ибо на Землю всегда действует сила, направленная к Солнцу. Если мы можем игнорировать все другие тела, считая их влияние очень незначительным, то нам больше ничего и не нужно; и мы можем воспользоваться доказательством Ньютона или более современным и простым доказательством.

Однако наши современные законы гравитации очень сложные. Прежде всего, мы утверждаем, что мировые линии, по которым тела свободно падают в пространстве-времени, являются геодезическими. Геометрия пространства определяется тензором массы-энергии, и т. д. Поэтому правомерен вопрос о том, как современная теория относи-

³ Это один из законов Кеплера.

⁴ При условии, что два тела — Солнце и Земля — исчерпывают собой вселенную. Если имеются другие силы, то закон Кеплера, *строго* говоря, нельзя считать правильным.

тельности объясняет закон Кеплера? Очень просто: *Законы Кеплера истинны, потому что законы Ньютона приблизительно истинны*. Если бы вместо этого аргумента мы попытались вывести законы Кеплера из уравнений гравитационного поля, то это было бы почти так же (но не совсем) нелепо, как и попытка дедуцировать из данных о местоположении и скорости отдельных атомов тот факт, что колышек проходит через одно отверстие и не проходит через другое.

Отсюда я делаю следующее философское заключение: *законы Ньютона обладают определенным рода реальностью в нашем мире, даже если они не истинны*. Дело в том, что они нужны для объяснения законов Кеплера. Я могу сделать это утверждение, по крайней мере, правдоподобным в методологическом плане. Алан Гарфинкель как-то заметил, что *хорошее объяснение остается инвариантным при небольших изменениях (perturbations) посылок*. Выведение законов Кеплера из уравнений гравитационного поля сталкивается с той проблемой, что, выполни мы сегодня это выведение, вполне может оказаться, что завтра эти уравнения гравитационного поля станут другими. Если же в качестве объяснения мы указываем, что, какими бы ни были наши уравнения, из них в первом приближении вытекают уравнения Ньютона, то это объяснение остается инвариантным даже при средних, а то и больших, изменениях посылок. Можно сказать, что каждое объяснение законов Кеплера «проходит» через законы Ньютона.

Вернемся, однако, к философии сознания. Если исходить из того, что атомная структура пронизывает всю материю, что имеется квантование и т. д., то, на первый взгляд, может показаться, что *непрерывность* не имеет никакого отношения к функционированию нашего мозга. Разве он не должен быть дискретным? Согласно физике, глубинный уровень материи дискретен.

Этот аргумент сталкивается с двумя проблемами. Первая состоит в том, что даже квантовая механика предполагает и длительность, и дискретность. Однако предположим, что квантовая механика полностью — теория дискретности.

Другая проблема заключается в том, что будь этот аргумент правилен, он доказывал бы, что при конструировании самолетных крыльев (по крайней мере тех, что предназначены для скоростей ниже сверхзвуковых) нельзя использовать в качестве модели воздуха такую непрерывную структуру, как жидкость. Отметим два момента: прерывистая, дискретная структура может аппроксимизировать непрерывную структуру. Дискретность может не иметь значе-

ния, как, например, в случае доски и колышка. Совершенно неважно, что колышек и доска — это твердые тела, не имеющие непрерывной структуры. Можно сказать, что доска и колышек — только приближенные копии абсолютно твердых непрерывных тел. Если погрешность аппроксимации неважна для этого уровня описания так, что с того? Дело ведь не только в том, что дискретные системы могут только аппроксимизировать непрерывные системы, но и в том, что система может вести себя так, как она ведет, *потому что* так же вела бы себя непрерывная система, для которой первая система выступает лишь аппроксимацией.

Наш мир — не ньютоновский. Совершенно четко. Закон Кеплера истинен, потому что система Солнце-Земля — это аппроксимация ньютоновской системы. И на этом уровне погрешность аппроксимации не играет никакой роли.

Предложенная аналогия не совершенна, поскольку физиков интересуют законы, для которых погрешность аппроксимации имеет значение. Думаю, для психологии эта аналогия более адекватна, ибо непрерывные модели (например, модель Халла для механического заучивания на основе потенциала непрерывности) могли бы быть вполне правильными независимо от того, какова глубинная структура мозга. Из факта существования нейронов нельзя сделать вывод о том, что цифровая модель должна быть правильной моделью. Мозг может функционировать так, как он функционирует, потому что он выступает аппроксимацией некоторой системы, законы которой лучше всего формулировать в терминах математики непрерывности. Более того, на уровне психологии погрешности этой аппроксимации не имеют значения.

То, что я сказал в отношении непрерывности, верно и в отношении других вещей. Вернемся к примеру о людях, имеющих душу, и людях, имеющих мозг, и к вопросу об изоморфизме между душами из одного мира и мозгами из другого. Возражение звучало так: если существует функциональный изоморфизм между душой и мозгом, то не слишком ли проста душа? Ответ — «нет». Ибо мозг может быть бесконечно сложно устроенным. Такая система как мозг, обладающая многими степенями свободы, может воспроизводить любую структуру с точностью в пределах, значимых для психологической теории. Возможно, что глубинная, так сказать, физика души довольно отличается от глубинной физики мозга, но на интересующем нас уровне, уровне функциональной организации, им обоим могло бы соответствовать одно и то же описание. Формально это описание может оказаться несовместимым с реальной физикой мозга — так

же, как *формально несовместимо с реальной структурой воздуха* описание воздуха, обтекающего корпус самолета, как некоторой несжимаемой сплошной жидкости.

В заключение я хотел бы отметить, что все эти примеры подтверждают ту идею, что субстанция, из которой мы состоим, не накладывает никаких первопорядковых (first order) ограничений на нашу форму. На деле нас интересует, как говорил Аристотель ⁶, не материя, а форма. Вопрос не в том, что есть материя, а в том, *что есть наша интеллектуальная форма?* И чем бы ни была наша субстанция — нечто духовное, материя или швейцарский сыр, она не накладывает никаких существенных ограничений на наш ответ. Конечно, она может накладывать ограничения более высокого порядка. Возможно, для объяснения незначительных явлений (affects) придется обратиться к реальной физике мозга. Но пока мы не достигли уровня *идеализированного* описания функциональной организации мозга, разговор о важности небольших пертурбаций еще преждевременен. Мой вывод: мы имеем то, что всегда хотели иметь — автономную ментальную жизнь. И для этого нам не требуется ни тайна, ни духовное начало, ни *жизненная сила*.

⁶ Так, Аристотель говорит: «... не следует спрашивать, есть ли душа и тело нечто единое, как не следует это спрашивать ни относительно воска и отпечатка на нем, ни вообще относительно любой материи и того, материя чего она есть» (412в 6-9). *Аристотель. О душе* // Аристотель. Соч. в 4-х томах. Т. 1. М., 1976, стр. 395.

1. ЧТО ТАКОЕ ЛОГИКА

Мы хотели бы начать с вопроса о том, *что есть* логика, а затем попытаться понять, почему должны существовать философские проблемы логики. Мы могли бы попытаться разобраться в том, что такое логика, исследуя различные определения «логики», но лучше начать с другого. Имеющиеся определения «логики» тем или иным способом сочетают в себе порочный круг и неточность. Вместо этого мы обратимся к самой логике.

Если мы обратимся к самой логике, то прежде всего заметим, что логика, подобно любой другой науке, подвергается изменениям, порой очень быстрым. В разные века у логиков были очень разные представления о границах их предмета, надлежащих методах исследования и т. д. Сегодня область логики определяют гораздо шире, чем когда-либо, ибо логика, как считают некоторые исследователи, включила в себя всю чистую математику. К тому же практически все методы, используемые сегодня в логическом исследовании, являются математическими. Другие аспекты логики, видимо, изменяются не столь значительно. Полученные однажды логические результаты, кажется, навсегда приняты как правильные, то есть, логика изменяется но не в том смысле, что мы в разные века принимаем несовместимые логические принципы, а в том, что стиль и система обозначений, используемые нами при формулировке логических принципов, значительно варьируют, а область, отведенная для логики, становится все шире и шире.

Поэтому, представляется разумным начать с рассмотрения некоторых принципов, которые логики приняли практически с самого начала. Один из этих принципов утверждает корректность (validity) ¹ следующего вывода:

* Putnam H. Philosophy of Logic. N. Y., Harper and Row, 1971.

¹ Английский термин «validity» по-разному переводится на русский язык в зависимости от того, выражает ли он свойства утверждений, рассуждений, выводов, формул, логических схем и т. д. (как, например, «общезначимость», «формальная истинность», «достоверность», «обоснованность» и т. д.). Патнэм в своей статье применяет термин «validity» ко всем перечисленным «абстрактным объектам» и, более того, делает выражаемое им свойство предметом своего рассмотрения. Именно поэтому мы выбрали в качестве перевода термин «корректность», который применим ко всем перечисленным случаям, хотя и достаточно неопределен. — *Прим. ред.*

- (1) $\frac{\text{Все } S \text{ есть } M}{\text{Все } M \text{ есть } P}$
 (следовательно) Все S есть P

Другим принципом является закон тождества:

- (2) x тождествен x .

Третий принцип говорит о *противоречивости* следующего утверждения:

- (3) p и не— p .

И, наконец, четвертый принцип утверждает корректность:

- (4) p или не— p .

Рассмотрим все эти принципы по порядку. Традиционно вывод (1) считается корректным при всех *терминах* S , M , P . Но что такое *термин*? В работах современных логиков обычно говорится, что вывод (1) корректен, независимо от того, какие *классы* можно приписать буквам S , M , P в качестве их экстенционалов. Тогда вывод (1) просто означает, что если класс S является подклассом класса M , а M в свою очередь является подклассом класса P , то S является подклассом P . Короче говоря, вывод (1) в его современной интерпретации просто выражает транзитивность отношения «быть подклассом». Это сильно отличается от того, что имели в виду логики прошлого, говоря о законах мышления и «терминах». Здесь мы сталкиваемся с обескураживающей особенностью науки логики: даже когда принцип может казаться неизменным на протяжении веков, например, мы по-прежнему пишем:

- Все S есть M
 Все M есть P
-
- (следовательно) Все S есть P

— *интерпретация* «неизменной» истины, на самом деле, значительно изменяется. Более того, остается спорной его «правильная» интерпретация.

Принцип (2) — еще один принцип, правильная интерпретация которого вызывает споры. Согласно интерпретации, предпочитаемой

большинством логиков (включая и автора статьи), принцип (2) утверждает рефлексивность отношения тождества: все находится в этом отношении (обозначаемом сегодня как «=») к самому себе. Однако некоторым философам крайне не нравится идея считать «=» отношением. «Разве можно видеть в *отношении*, — говорят они, — нечто иное, чем то, что одна вещь может иметь к *другой* вещи?» Поскольку ничто не имеет тождества с *другой* вещью, то, заключают они, символ «=» может обозначать все, что угодно, но только не отношение.

Наконец, в отношении принципов (3) и (4) встает проблема: что обозначает *p*? Некоторые философы, например, убеждены, что в (4) *p* обозначает любое *предложение*, в то время как другие философы, включая и автора статьи, находят нелепой теорию, согласно которой предметом логики являются *предложения*.

Однако, все эти разногласия относительно тонкостей не должны скрывать от нас того факта, что существует значительное согласие среди логиков — даже среди логиков разных веков. Все логики согласны, например, что из посылок:

Все люди смертны
Все смертные неудовлетворены

корректно сделать вывод, что:

Все люди неудовлетворены,

даже если они иногда несогласны относительно точной *формулировки* общего принципа, лежащего в основе этого вывода. Сходным образом, все логики согласны, что если существует такая вещь, как Эйфелева башня, то

Эйфелева башня тождественна Эйфелевой башне,

и все логики согласны, что (если существует такая вещь, как «земля», то)

Земля является круглой или земля не является круглой,

даже если они и в этих случаях расходятся относительно формулировки соответствующего принципа. Таким образом, в логике имеется некоторая доля «перманентной теории», но она не слишком помога-

ет, по крайней мере, при формулировке точных и универсально приемлемых общих принципов.

II. ВОПРОС О НОМИНАЛИЗМЕ И РЕАЛИЗМЕ

Теперь уже ясно, что существуют философские проблемы логики, а также ясна, по крайней мере, одна из причин этого, связанная с трудностью получения любой универсально приемлемой формулировки общих принципов, которую так или иначе признали бы все логики. Если мы исследуем эту трудность, то станут яснее и другие философские проблемы логики.

Философов и логиков, считающих классы, числа и сходные «математические сущности чем-то выдуманным, обычно называют «номиналистами». Номиналист не склонен утверждать, что:

(А) «Для всех классов S , M , P : если все S есть M , и все M есть P , то все S есть P ».

Скорее, он скажет:

(В) «Нижеследующее превратится в истинное предложение, какие бы слова или фразы соответствующего вида, мы ни подставили вместо букв S , M , P : “Если все S есть M , и все M есть P , то все S есть P »».

Причина, если и не убедительна, то хотя бы ясна: номиналист, по сути, не верит в существование классов; поэтому он избегает использовать формулировку (А). В отличие от классов, «предложения» и «слова» представляются ему более «реальными», и поэтому он применяет формулировку (В).

Стало быть, отчасти разногласия по поводу «правильной» формулировки наиболее общих принципов логики — это лишь отражение философских разногласий по поводу существования или несуществования таких «математических сущностей», как классы.

Однако, независимо от преимуществ той или иной позиции в вопросе «номинализм или реализм», ясно, что на деле формулировка (В) не может быть предпочтительнее формулировки (А). Ибо что имеется в виду под «словом или фразой соответствующего вида» в формулировке (В)? Даже если мы отставим в сторону вопрос о том, что же собственно определяет «соответствующий вид» слова или фразы, мы неизбежно окажемся перед тем фактом, что в этой формулировке подразумеваются все возможные слова или фразы того

или иного рода, а эти *возможные слова и фразы* сами не более «реальны», чем классы.

Иногда от этого вопроса уходят тем или иным способом. Один из таких способов — охарактеризовать все соответствующие «фразы», подставляемые вместо S , M и P как «одноместные предикаты» некоторого «формализованного языка». Формализованный язык задается полным описанием грамматики вместе с указанием значений исходных выражений. Какие выражения в таком языке являются одноместными предикатами (т. е. именами классов, хотя номиналист никогда не стал бы их так называть) — устанавливается правилом формальной грамматики. По существу, если дан формализованный язык L , то класс допустимых подстановок вместо псевдобукв (dummy letters) S , M , P в выражении:

(5) Если все S есть M , и все M есть P , то все S есть P

можно определить с высокой точностью, а, стало быть, можно чисто механическим способом (например, с помощью компьютера) установить, является ли определенная цепочка букв «результатом подстановки» (substitution-instance) в вывод (1).

Это почти разрешает сомнения номиналиста, ибо утверждая корректность вывода (5), мы в этом случае, казалось бы, вовсе не говорим о «классах», а просто утверждаем, что истинны все «результаты подстановок в вывод (5)» (в некотором определенном языке L); т. е. истинны все *цепочки букв*, которые удовлетворяют некоторому формальному критерию (и, стало быть, являются результатами подстановок в вывод (5) в формализованном языке L). И, конечно же, «цепочки букв» совершенно реальны — разве нет?

К сожалению, для номиналиста сразу же возникают серьезные трудности. Под логической *схемой* обычно понимают выражение типа (5), составленное из «псевдобукв», таких, как S , M , P , и логических слов «если—то», «все», «некоторые», «или», «не», «тождественно», «есть» и т. п. Такие схемы используются логиками со времен Аристотеля и до наших дней для формулировки логических принципов (если Аристотель сосредоточил свое внимание на очень ограниченном классе схем, то современные логики исследуют все возможные схемы описанного типа). Схема может быть, подобно выражению (5), «корректной», т. е. она может выражать «правильный» логический принцип (что такое правильность или корректность нам еще предстоит рассмотреть), или она может быть «некорректной». Например, утверждение:

Если некоторые S есть P , то все S есть P

представляет собой пример некорректной схемы — схемы, которая не выражает правильный логический принцип. Уже античные и средневековые логики дали классификацию огромного количества схем как корректных или некорректных.

Очевидно, что определение корректности ставит серьезные философские проблемы. Но определение корректности, которое мы только что приписали номиналисту и согласно которому схема S корректна только в том случае, если истинны все результаты подстановок в нее *в некотором конкретном формализованном языке L* , — это определение уже с первого взгляда выглядит неудовлетворительным. Ибо, когда я говорю, что вывод (5) корректен, я считаю его корректным *независимо от того, какие имена классов можно подставить вместо S , M , P* . Если бы некоторый формализованный язык L содержал имена для всех классов вещей, которые можно составить, то это было бы равносильно утверждению, что «все результаты подстановок в (5) в языке L истинны». Но согласно одной из теорем теории множеств *никакой язык L не может* содержать имен для *всех* возможных совокупностей вещей, по крайней мере, если число вещей бесконечно.

Другими словами, приняв предложение номиналиста, мы получили не *единственное* понятие корректности, а бесконечное множество таких понятий: корректность в L_1 , корректность в L_2 , корректность в L_3 , ..., где каждое из этих понятий означает просто «истинность всех результатов подстановок» в соответствующем языке L_i .

Мы могли бы пытаться избежать этого, сказав, что схема S корректна только в том случае, если истинны все результаты подстановок в каждом языке L ; но тогда нам нужно понятие *всех возможных формализованных языков* — понятие, гораздо менее «реальное», чем понятие «класс».

Во-вторых, предложенное номиналистическое определение корректности предполагает понятие «истины», а это понятие для номиналистов проблематично. Обычно мы не считаем «истинными» или «ложными» материальные объекты, например, цепочки реально написанных букв (воспринимаемых как небольшие чернильные пятна на бумаге); скорее истинным или ложным является *то, что цепочки букв выражают*. Но именно значение цепочки букв или то, что эта цепочка букв «выражает», и есть тот род сущностей, от которых номиналист хочет избавиться.

В-третьих, когда мы говорим обо *всех* подстановках в схему (5), даже применительно к одному конкретному языку L , мы имеем в виду все *возможные* подстановки, а не только «существующие» в номиналистическом смысле (как небольшие чернильные пятна на бумаге). Когда мы утверждаем, что схема (5) корректна, мы не просто имеем в виду, что *истинны те варианты этой схемы, которые случилось записать*, ибо, возможно, что имеется ложный результат подстановки в схему (5), который просто еще не был записан. Но *возможные* подстановки в схему (5) или *возможные* цепочки букв являются не более реальными физическими объектами, чем классы.

Думается, одну из проблем можно разрешить с помощью нижеследующих соображений. Нет оснований при формулировке логических принципов в большей мере избегать ссылок на «нефизические сущности», чем в научном дискурсе в целом. Ссылки на классы вещей, а не просто на вещи, — это обычная и полезная манера говорить. Если номиналист желает, чтобы мы отказались от нее, он должен снабдить нас альтернативной манерой, столь же хорошо работающей не только в чистой логике, но и в таких эмпирических науках, как физика (которая изобилует ссылками на такие «нефизические» сущности как векторы состояний, гамильтонианы, гильбертово пространство и т. д.). Если он когда-нибудь преуспеет в этом деле, это повлечет изменения в формулировке не только логических, но и всех научных принципов. Между тем, нет оснований отказываться от формулировок, подобных (А), ввиду серьезных проблем, порождаемых такими формулировками, как (В). (И, как мы только что видели, формулировка (В), не будучи адекватной, вдобавок к тому же не является и по-настоящему номиналистической.)

Иными словами, тот факт, что формулировка (А) «вызывает возражения» с номиналистической точки зрения, в действительности, составляет трудность не для науки логики, а для философии номинализма. Логика, как и любая другая наука, вовсе не обязана подстраивать используемую в ней манеру говорить к философским требованиям номинализма; скорее номиналист должен обеспечить удовлетворительную переинтерпретацию таких утверждений, как (5), равно как и любых других высказываний логиков (физиков, биологов, и даже простых людей с улицы).

Даже если мы отбросим номинализм как требование очистить здесь и сейчас наш научный язык от всех ссылок на «нефизические сущности», мы не должны, тем не менее, отбрасывать номинализм как философию. Те, кто полагает, что на самом деле ничто не соответствует таким понятиям, как класс, число, возможная цепочка

букв, а если и соответствует, то это свидетельствует о в высшей степени производной манере говорить об обычных материальных объектах, могут и дальше придерживаться своей точки зрения, а наше нежелание изменить обычный научный язык согласно их требованиям ни в коей мере не является нежеланием обсуждать философские проблемы, поднятые их позицией. К этому мы теперь и перейдем.

Можно начать с рассмотрения различных трудностей, только что обнаруженных нами в связи с формулировкой (В), и установить, какие возражения может выдвинуть номиналист в ответ на эти трудности.

Сначала одно-два общих замечания. Нелсон Гудмен, наиболее известный философ номинализма, никогда не принимал определения корректности как «истинности всех подстановок». (Это идет от Хьюго Леблана и Ричарда Мартина.) Однако, Гудмен никогда и не занимался проблемой определения логической корректности, поэтому я беру на себя смелость обсудить ту квази-номиналистическую попытку, о которой знаю. Во-вторых, Гудмен отрицает, что номинализм — это требование ограничиться «физическими» сущностями. Хотя, быть может, утверждение о том, что реальными являются только физические сущности (или только «ментальные конкретные сущности», согласно идеалистическому варианту номинализма; или только ментальные конкретные сущности и физические вещи, согласно дуалистическому варианту), и не выражает позиции Гудмена, но именно эту точку зрения большинство людей связывают с «номинализмом», и, думается, что у номиналистов нет особого повода отказываться от этой точки зрения. (Различие между требованием ограничиться «физическими сущностями» и требованием ограничиться «ментальными конкретными сущностями» или «физическими вещами и ментальными конкретными сущностями» не будет обсуждаться в настоящей статье, поскольку оно не затрагивает существенным образом философию логики.)

Первый аргумент, выдвинутый нами против формулировки (В), состоял в том, что эта формулировка, по сути, подменяет наше интуитивное понятие корректности множеством таких понятий, и это множество содержит столько понятий, сколько возможно формализованных языков. Некоторые логики пытались устранить эту трудность, используя следующий прием: пусть L_0 — формализованный язык, достаточно богатый, чтобы на нем можно было формулировать утверждения о положительных целых числах и выражать понятия « x есть сумма y и z » и « x есть произведение y и z ». Пусть L_i — любой

другой формализованный язык. И, наконец, пусть S — схема, обладающая тем свойством, что все результаты подстановок в нее в языке L_0 истинны (назовем это свойство свойством быть «корректным— $v-L_0$ » и, аналогичным образом, назовем схему «корректной— $v-L_1$ », если все результаты подстановок в нее в языке L_1 являются истинными). Тогда можно доказать, что схема S обладает и тем свойством, что все результаты подстановок в нее в языке L_1 также истинны, и это доказательство может быть формализовано в любом языке, достаточно богатом, чтобы содержать понятия «истины в L_0 » и «истины в L_1 ». Другими словами, *если схема корректна— $v-L_0$, то она также корректна— $v-L_1$* . Поэтому, предлагают эти логики, «корректность» можно просто определить как *корректное— $v-L_0$* . Если S корректна, то отсюда следует — не по определению, а в силу только что упомянутой метаматематической теоремы, — что истинны и все результаты подстановок в нее в языке L_1 независимо от того, каков язык L_1 . Поэтому, «корректность» схемы дает нам основание считать эту схему истинной при любых произвольных подстановках в нее (как и должно быть, согласно нашему интуитивному понятию).

В ответ на это, можно было бы возразить, что когда я говорю « S корректна» — это непосредственно означает, что каждый результат подстановки в схему S (в каждом формализованном языке) является истинным. Согласно предложенному определению корректности, говоря « S корректна», я имею в виду лишь то, что все результаты подстановок в схему S в языке L_0 являются истинными; поэтому то, что результаты подстановок в схему S в любом языке истинны, — это просто математический факт, а не часть подразумеваемого мной смысла. Таким образом, предложенное определение корректности совершенно не отражает содержания интуитивного понятия, даже если и совпадает с этим интуитивным понятием по объему.

Однако это возражение необязательно разрушительно, поскольку логик-номиналист может в ответ сказать, что он вовсе не стремится отразить «интуитивное» понятие; достаточно, если он обеспечивает нас философски приемлемым (с его точки зрения) понятием, ибо это соответствует поставленной им задаче.

Пусть это так, но факт остается фактом: сам язык L_0 требует ссылок на «математические сущности» (в частности, числа), и для доказательства утверждения, что «если S корректна— $v-L_0$, то S корректна— $v-L_1$ » необходимы ссылки на произвольно взятые выражения L_1 (т. е. на все возможные выражения L_1). Таким образом, ни язык L_0 , ни только что упомянутая метаматематическая теорема в

действительности не годятся *строгому* номиналисту, т. е. тому, кто избегает *любых* разговоров о «математических сущностях».

Второй сформулированный нами аргумент состоял в том, что и понятие «истина» не подходит для номиналиста. Однако, это утверждение весьма спорно.

Если говорить в двух словах, то мы утверждали, что выражение «истинно» не имеет смысла в применении к физическому объекту, даже если этот физический объект — написанное предложение; истинным или ложным является не само написанное предложение, а *то, что оно говорит*. И те вещи, которые выражают предложения, в отличие от самих предложений (записей), не являются физическими объектами.

Для номиналиста в данном случае будет естественным, я полагаю, провести различие между

(6) S истинно

и

(7) S , как его понимает Оскар во время t , истинно

Если S представляет собой физический объект (например, написанное предложение), то предложение (6) действительно имеет много смысла, если только оно не служит кратким выражением некоторого факта, подобного (7). Но предложение (7) выражает весьма вероятное отношение, которое может иметь место или не иметь места между данным написанным предложением, человеческим организмом и временем. (Я не буду рассматривать, как номиналисту следует трактовать ссылки на «время»; возможно, он должен отождествить «время» с соответствующим трехмерным поперечным сечением в четырехмерном пространственно-временном универсуме.) Разве у номиналиста нет возможности утверждать, что некоторые *предложения истинны* в том смысле, что сами находятся в отношении, представленном в (7), к соответствующим организмам в соответствующее время? Тогда, учитывая сложный характер этого отношения, бремя доказательства ляжет на реалиста, который должен будет показать, что это отношение по сути предполагает существование не физических сущностей таких, как высказывания, значения и т. п.

В другой формулировке этот второй аргумент принимает вид «апелляции (арреал) к обыденному языку». Так, утверждается, что

(8) Джон сделал истинное утверждение

— это совершенно правильный случай употребления «обыденного языка» в определенных легко представимых ситуациях. Стало быть, имеются две возможности: (а) из (8) следует, что утверждения существуют (как нефизические сущности); и (б) из (8) этого не следует. В случае (б) нет проблем; мы можем и дальше рассуждать об «утверждениях» (а, коли на то пошло, и о «классах», «числах» и т. п.), поскольку, как все согласны, подобные рассуждения не предполагают существования утверждений (чисел и классов) как нефизических сущностей. Тогда номинализм бесполезен, поскольку те лингвистические формы, от которых он стремится избавиться, безвредны с философской точки зрения. Что касается случая (а), то поскольку (8) истинно и предполагает существование нефизических сущностей, отсюда следует, что эти нефизические сущности существуют! Поэтому номинализм ложен! Таким образом, номинализм или бесполезен, или ложен.

В ответ номиналист может возразить, что он стремится найти лишь «функцию-перевод», позволяющую заменить такие предложения как (8) предложениями, которые даже *по виду* не предполагают существования нефизических сущностей. В результате, полагает он, мы будем обеспечены терминологией, менее запутанной в концептуальном отношении и более соответствующей природе вещей, чем та, которой мы обычно пользуемся. Безусловно, такие предложения, как (8), «философски безвредны» при их правильном истолковании; но проблема состоит в том, чтобы прояснить это правильное истолкование.

Для пушей убедительности номиналист может добавить, что, с его точки зрения, функция-перевод необязательно должна сохранять синонимию. Достаточно, если истолкование предложений, подобных (8) в соответствии с моделью номиналистического перевода будет обеспечивать бóльшую ясность.

Таким образом, тот факт, что в «обыденном языке» слова «истинное» и «ложное» обычно применяются к «утверждениям», не убеждает номиналиста *ни в том*, что утверждения действительно существуют как нефизические сущности, *ни в том*, что отход от обыденного языка (и использование таких предложений, как (7)), есть интеллектуальное прегрешение.

Наконец, есть «аргумент», согласно которому предложение (7) означает следующее: имеется *утверждение*, которое схема *S* «выражает» для Оскара во время *t*, и это *утверждение* истинно. Это означает, что предложение (7) содержит скрытую ссылку на нефизи-

ческую сущность (на то, что S «выражает») и, следовательно, не является «на самом деле» номиналистическим.

Этот аргумент сводится либо к тому, что рассмотренному аргументу с апелляцией к обыденному языку, либо к простому требованию считать, что *на самом деле* только *утверждения* (понимаемые как нефизические сущности, выражаемые предложениями) могут быть «истинными» или «ложными». Поскольку именно это требование и составляет предмет нашего обсуждения, этот аргумент ничего не доказывает.

Итак, все аргументы, доказывающие непригодность понятия истины для номиналиста, выглядят несостоятельными. С другой стороны, отсюда вовсе не следует, что номиналист *имеет право* использовать это понятие. Истина (или триадическое отношение между написанными предложениями, человеческими организмами и временем, представленное в (7)) едва ли такая же примитивная вещь, как понятие «желтый», поэтому номиналист должен дать *некоторое* разъяснение понятия истины, которое было бы непротиворечивым образом сформулировано в категориях его метафизики. Если он не сможет дать нам такого разъяснения (а что он тогда имеет?), его право на использование данного понятия становится сомнительным.

Прежде чем читатель (или номиналист) успеет воскликнуть: «Tu quoque!» («И ты тоже!»), напомним ему следующие факты: «интуитивное» понятие истины представляется противоречивым (см. в связи с этим хорошо известные логические антиномии); но если задан какой-нибудь формализованный язык L , то имеется и предикат «истинно—в— L », который можно использовать во всех научных целях вместо интуитивного понятия истины (если рассматриваемые утверждения сформулированы в языке L), и *этому* предикату можно дать *точное* определение, используя лишь выражения языка L и теорию множеств². Это не совсем удовлетворительное решение, поскольку предпочтительнее было бы иметь единственный предикат «истинно», а не бесконечное множество предикатов «истинно—в— L_1 », «истинно—в— L_2 » и т. д., однако оно вполне сносно, а антиномии дают серьезные основания сомневаться в том, что вообще возможно непротиворечивое понятие истины, применимое ко *всем* языкам и удовлетворяющее интуитивным требованиям. Таким образом, реалист может не разъяснять интуитивное понятие истины, а выработать целый ряд альтернативных понятий и использовать их во

² Это было показано А. Тарским. Достаточно популярное изложение этого вопроса см. в статье: *Tarski A. The semantic conception of truth // Feigl H., Sellars W. (eds.) Readings in the Philosophical Analysis, N. Y., p. 52—84.*

всех научных контекстах, где требуется предикат «истинно», давая им строгое определение. Но номиналист — по крайней мере, сегодня — не может сделать даже этого.

Согласно нашему третьему аргументу, ссылаясь на *все* предложения формализованного языка (или даже на все результаты подстановок в некоторую фиксированную схему), мы ссылаемся не на «записи» (поскольку едва ли можно предположить, что все бесконечное множество предложений некоторого формализованного языка где-то реально записано), а на абстрактные сущности, «возможные записи», или, как говорят некоторые авторы, на «типы» или формальные свойства (shape-properties), представленные этими записями. (Предполагается, что эти типы «существуют» независимо от того, представлены они реально существующими записями или нет; иными словами, они также являются нефизическими сущностями.) Когда мы говорим «все результаты подстановок в схему (S) истинны», мы имеем в виду *даже те результаты подстановок, которые еще никто реально не записал*. Таким образом, эти «результаты подстановок», особенно «потенциальные», являются не в большей мере «физическими», чем классы. Насколько мне известно, на этот аргумент нет ответа, достойного рассмотрения.

Итак, наше рассмотрение трех указанных аргументов не изменило нашего вывода о том, что формулировка (B) не является номиналистической. Однако, мы увидели, что чем больше мы углубляемся в первые два аргумента, тем сложнее (и техничнее) они становятся.

Мы можем подытожить этот раздел, сказав, что на сегодняшний день ссылки на «классы» или на нечто столь же «нефизическое» неустранимы из науки логики. Понятие логической «корректности», на котором основывается вся наука, нельзя, по крайней мере сегодня, удовлетворительным образом объяснить в чисто номиналистических терминах.

III. ВОПРОС О НОМИНАЛИЗМЕ И РЕАЛИЗМЕ И ЛОГИКА

Антитеза номинализм—реализм существует с давних пор, и было бы интересно проследить, как она оказалась связанной с философией логики. В элементарной логике со времен Аристотеля были сформулированы такие принципы, как (2), (4), (5), был составлен список схем корректного вывода, включающих (1), и было заявлено о противоречивости таких форм, как (3). По сравнению с логикой

Аристотеля, современная «теория квантификации» — так называется соответствующий раздел современной логики, — или «первопорядковая логика с тождеством», имеет гораздо более широкую область применения, но весьма сходные проблемы.

Исходные символы таковы:

(i) « Px » означает « x есть P », и, сходным образом, « Pxy » означает « x находится в отношении P к y », а « $Pxyz$ » — « x , y , z находятся в отношении P » и т. п.;

(ii) « (x) » (читай: «для каждого x ») означает, что каждый объект (entity) x удовлетворяет некоторому условию; т. е. « $(x) Px$ » означает «каждый объект x есть P ».

(iii) « (Ex) » (читай: «существует x такой, что») означает, что некоторый (по крайней мере один) объект x удовлетворяет какому-то условию; т. е. « $(Ex) Px$ » означает «существует объект x , который есть P ».

(iv) « $=$ » (читай: «тождественно») обозначает тождество, т. е. « $x = y$ » означает « x тождествен (есть тот же самый объект, что и) y ».

(v) « \vee » означает «и», « \vee » означает «или», « \neg » означает «не», например, « $(Px \vee \neg Qx) \cdot Rx$ » означает «или x есть P или x есть не- Q ; и x есть R ».

Кроме того, символы \supset (читай: «если... то») и \equiv (читай: «если и только если») вводятся с помощью определений: « $Px \supset Qx$ » («Если Px , то Qx ») служит краткой записью для « $\neg (Px \cdot \neg Qx)$ », а « $Px \equiv Qx$ » служит краткой записью для « $(Px \supset Qx) \cdot (Qx \supset Px)$ ».

В этой системе обозначений мы можем записать все сформулированные Аристотелем принципы. Например, принцип (5) принимает вид:

$$(5') ((x) (Sx \supset Mx) \cdot (x) (Mx \supset Px)) \supset (x) (Sx \supset Px).$$

Принимая во внимание весь класс схем, которые можно записать с помощью этой системы обозначений, мы должны отметить и потенциальные логические принципы, которые Аристотель никогда не анализировал, поскольку его интересовали лишь выводы, где каждая посылка содержит только два имени класса.

Однако, более важно то, что с помощью современной системы обозначений мы можем анализировать выводы, по существу предполагающие двух-и-более-местные *отношения*; именно из-за отсутствия аналога отношений, логика, изучаемая до конца XIX века, имела довольно тривиальный характер, а потому вся традиционная логика — от Аристотеля и до Буля включительно, чьи работы крайне важны для последующих логических исследований, — оказалась совершенно непригодной для анализа дедуктивных рассуждений в их более сложных формах.

Во многих своих логических и философских работах Куайн подчеркивал, что теория квантификации не утверждает, например, истинности формулировки (A), представленной в предыдущем разделе. С точки зрения Куайна, строя систему, где, скажем, схема (5') является одной из теорем, логик не утверждает тем самым истинность принципа (A). Напротив, в схемах (5) и (5') S, M, P — это «псевдобуквы» для *каких угодно предикатов*, а логик утверждает лишь, что *все результаты подстановок в схему (5) или (5') являются логическими истинами*. Согласно этой точке зрения, следующее утверждение является «логической истиной»:

(9) Если все вороны черные и все черные вещи поглощают свет, то все вороны поглощают свет.

Однако общий принцип (A):

Для всех классов S, M, P : если все S есть M , и все M есть P , то все S есть P —

с точки зрения Куайна не *логическая*, а *математическая* истина.

Меня не слишком заботит то, где проводится граница между логикой и математикой, но это конкретное решение Куайна не представляется мне убедительным.

У меня есть два принципиальных довода. Во-первых, логическая традиция свидетельствует против Куайна: ибо с самого начала в задачу логиков входила формулировка таких общих принципов, как (A), а не «вычленение» таких истин, как (9), из числа других истин. Во-вторых, не думаю, что все результаты подстановок в корректную схему являются «истинными»: некоторые очевидным образом бессмысленны. Например:

(10) Если все буджумы есть снарки, и все снарки есть эгеламфы, то все буджумы есть эгеламфы, —

не представляется мне истинным утверждением; оно имеет форму логически корректного утверждения, но, думаю, оно вообще не является утверждением — ни истинным, ни ложным. Действительно, если назвать предложение (10) истинным, это потребует пересмотра обычных логических правил. Согласно теореме стандартной логики, если утверждение, имеющее форму «если p и q , то r » истинно, то или p и q оба истинны и r истинно, или p истинно, q ложно и r истинно (или ложно), или p ложно, q истинно и r истинно (или ложно), или p и q оба ложны и r истинно (или ложно). Но в случае предложения (10) все три компонента, соответствующие p , q и r , не являются ни истинными, ни ложными.

Конечно, можно было бы расширить понятие истины и называть любое утверждение, имеющее форму логически корректного утверждения, истинным. Однако в этом случае предложение:

(11) Все буджумы есть снарки или неверно, что все буджумы есть снарки,

(имеющее форму $p \vee \neg p$), следовало бы считать истинным, а это представляется совершенно неверным, поскольку обычно предполагается, что если человек утверждает (11), то он обязан считать, что:

(12) Утверждение, что все буджумы есть снарки, или истинно, или ложно.

С моей точки зрения, логика, как таковая, не говорит нам, что (9) истинно: для знания того, что (9) истинно, я должен воспользоваться моим знанием логического принципа (А) и знанием того факта, что каждый из предикатов « x есть вóрон», « x черный» и « x поглощает свет» истинен относительно вещей определенного класса, а именно класса вóронов, класса черных вещей и класса вещей, поглощающих свет, соответственно. Даже это «знание» предполагает определенную идеализацию, связанную с тем, что здесь игнорируется тот факт, что некоторые из этих предикатов (в частности, черный) нечетко определены (не являются ни истинными, ни ложными) в некоторых случаях. Даже, если мы склонны совершать такую идеализацию, то и в этом случае знать, скажем, что « x есть вóрон» — это предикат, истинный (помимо возможных промежуточных случаев) относительно всех вещей определенного класса и ложный относительно всех вещей, не входящих в этот класс, значит знать немало о языке и мире. Если мы знаем, что предикат « x есть

вóрон» — достаточно хорошо определен, предикат « x красива» — нечетко определен, а « x есть снарк» — бессмысленен, то это наше знание никак не является *логическим*.

Итак, наше разногласие с Куайном состоит в том, что Куайн считает «логическими истинами» именно такие утверждения, как (9), а с моей точки зрения, каждое такое утверждение отражает сложную совокупность логического и нелогического знания. Однако, здесь не требуется, чтобы читатель согласился со мной, а не с Куайном. В целях настоящего исследования я настаиваю лишь на том, что решение называть такие утверждения, как (A), «принципами логики» является достаточно обоснованным и с исторической, и с концептуальной точки зрения. Конечно, здесь возможен выбор, но для нас важно то, что при одном достаточно естественном решении утверждения типа (A), содержащие явные ссылки на классы, составляют часть *логики*.

Рассмотренные нами логические схемы содержали (x) [для каждого индивидуального объекта x] и (Ex) [существует индивидуальный объект x такой, что], но не (F) и (EF). Так, при заданном «универсуме рассмотрения» мы можем утверждать, используя введенную систему обозначений, что некоторый элемент универсума есть P , записав это в виде (Ex) Px , но мы не можем утверждать, что существует множество или класс всех элементов, обладающих свойством P (в символическом виде: (EF) (x) ($Fx \equiv Px$)), потому что в нашей системе обозначений нет «(EF)».

Великие основатели современной логики — Фреге, а вслед за ним, и Бертран Рассел, без колебаний сочли такие выражения как (EF) частью логики и даже допустили в качестве части «логики» такие выражения как (EF^2), означающие для каждого класса классов, и (EF^3), означающие для каждого класса классов классов и т. п.

Я не думаю, что это было ошибкой. Быть может, их решение не было единственно возможным. Действительно, во введении ко второму изданию «*Principia Mathematica*» Рассел осмотрительно воздерживается от утверждения, что их решение единственно возможное, однако, оно представляло собой совершенно естественный выбор. Ответ на вопрос: «Где «провести границу» (если мы вообще хотим проводить эту границу) между логикой и теорией множеств (а, следовательно, и между логикой и математикой)?» — неизбежно будет носить произвольный характер.

Предположим, тем не менее, что мы решили провести границу на уровне логики «первого порядка» («теории квантификации») и считаем такие выражения как (EF), (EF^2) и т. д. принадлежащими к

«математике». Но по-прежнему остается проблема: *что утверждает логик*, строя системы, содержащие теоремы, подобные (5')? Возможно, конечно, он ничего и не утверждает, а просто конструирует неинтерпретированную формальную систему, но тогда он определенно не занимается логикой. Дело в том, что большинство логиков хотело бы утверждать следующее: теоремы системы являются корректными формулами. Если не явным образом, то хотя бы имплицитно, логик стремится делать утверждения, имеющие форму «то-то и то-то *корректно*»; т. е. утверждения типа (А). Поэтому даже логику первого порядка обычно понимают так, будто она «метатеория»; если вообще логик делает какие-то утверждения, когда записывает схемы, подобные (5'), то он делает утверждения о корректности, а это означает, что он неявным образом делает утверждения второго порядка: утверждать корректность схемы первого порядка (5') значит утверждать, что (S) (M) (P) (схема 5'), а это уже утверждение второго порядка.

Подведем итог. На мой взгляд, (а) довольно неоправданно утверждать, что логика «второго порядка» не является «логикой»; но (б) даже если мы действительно так думаем, вполне естественно считать, что записывая схемы первого порядка, мы неявно утверждаем их корректность, т. е. делаем утверждения второго порядка. В свете этого нетрудно понять, почему и каким образом традиционный вопрос о реализме и номинализме привлек огромное внимание философов логики: ибо если мы правы, то вполне естественно считать, что вся логика, включая и теорию квантификации, содержит ссылки на классы, т. е. на тот род сущностей, который хотел бы устранить номиналист.

IV. ЛОГИКА ИЛИ МАТЕМАТИКА

В свете наших предыдущих рассуждений понятно, почему крайне трудно избежать произвольности при проведении границы между логикой и математикой. По мнению некоторых мыслителей, эту границу следует провести между логикой первого порядка и логикой второго порядка. Однако, как мы только что видели, это имеет то неудобное следствие, что понятия корректности и импликации³ оказываются принадлежащими не к логике, а к математике. Фреге, Рассел и Уайтхед относили не только логику второго порядка, но даже и ло

³ Считается, что «*A* имплицирует *B*» только в том случае, если корректно условное высказывание ($A \supset B$), где *A* — антецедент, а *B* — консеквент. Короче говоря, «вывод — это корректное условное высказывание».

гику более высоких порядков (множеств множеств множеств... множеств индивидуальных объектов) к логике. Это решение равносильно утверждению, что не существует границы «между» математикой и логикой; математика составляет часть логики. Если кому-то по душе «промежуточная» позиция, то, возможно, рассматриваемую «границу» следует провести между логикой второго порядка и логикой третьего порядка. Однако, мы не будем особенно вдаваться в эту проблему. Обсуждаемые в этом параграфе философские вопросы затрагивают как философию математики, так и философию логики, и мы не будем пытаться развести эти две дисциплины порознь.

V. НЕАДЕКВАТНОСТЬ НОМИНАЛИСТИЧЕСКОГО ЯЗЫКА

Под «номиналистическим языком» понимается формализованный язык, в котором переменные определены, в некотором подходящем смысле, на множестве индивидуальных объектов, а предикатные символы обозначают прилагательные и глаголы, применяемые к этим индивидуальным объектам (например, «тяжелый», «больше чем», «быть частью»). Эти прилагательные и глаголы необязательно соответствуют наблюдаемым свойствам и отношениям; например, вполне допустим предикат «быть электроном», однако они не должны предполагать существования такие сущностей, как классы и числа.

Неоднократно отмечалось, что такой язык не годится для целей науки; если принять, что с философской точки зрения это единственный язык, который мы имеем право использовать, то это потребовало бы отказаться практически от всей математики. Но на деле ограничения номинализма разрушительны как для формальной, так и для эмпирической науки; мы должны были бы отказаться не только от «математики», но и от физики.

Чтобы понять это, рассмотрим наиболее известный пример физического закона: закон всемирного тяготения Ньютона. (Для настоящего обсуждения неважно, что этот закон не является, строго говоря, истинным; формулировка более сложных и действительно истинных законов потребовала бы гораздо более сложный математический аппарат.) Закон Ньютона, как всем известно, утверждает, что любое тело a действует на любое другое тело b с силой f_{ab} . Эта сила f_{ab} направлена в сторону a , и ее величина F определяется формулой:

$$F = \frac{g M_a M_b}{d^2}$$

где g — универсальная константа, M_a — масса тела a , M_b — масса тела b , а d — расстоянием между a и b .

Придерживаясь в данном случае «реалистической» философии физики, я буду исходить из того, что одной из основных целей физики является формулировка «истинных или приблизительно истинных» (по словам Ньютона) законов, а не просто установление связей между опытными данными или их предсказание. Также я буду предполагать, что приведенный выше закон правилен, хотя мы знаем сегодня, что он только приблизительно выражает гораздо более сложный закон. Оба эти допущения должны быть приемлемыми для номиналиста. Мне кажется, в глубине души номиналисты должны быть материалистами, ибо в противном случае все их усилия непонятны. А материалисту не следует сомневаться в том, что материя подчиняется некоторым объективным законам и цель науки — сформулировать эти законы. Мы допускаем, что закон Ньютона, строго говоря, истинен лишь для того, чтобы иметь перед собой пример физического закона, формулируемого в математическом виде (потому он и невыразим в номиналистическом языке) и понятного большинству людей, чего, к сожалению, не скажешь о многих более сложных физических законах.

Итак, главное в нашем примере то, что закон Ньютона имеет содержание, которое, с одной стороны, совершенно ясно (закон гласит, что гравитационная «сила тяги» прямо пропорциональна массам тел и обратно пропорциональна квадрату расстояния между ними), а с другой стороны — выходит за пределы того, что можно выразить в номиналистическом языке. Даже если бы мир был проще, чем он есть, гравитация была бы единственной силой, а закон Ньютона выполнялся точно, все равно было бы невозможно «заниматься» физикой, используя номиналистический язык.

Однако как мы можем быть в этом уверены? Даже если ни один номиналист не предложил еще способа «перевода» утверждений, подобных закону Ньютона, на номиналистический язык, как мы можем быть уверены в том, что такого способа не существует?

Рассмотрим не только сам закон гравитации, но и его очевидные предпосылки. В первую очередь, закон предполагает существование сил, расстояний и масс — возможно не как реальных сущностей, но как вещей, которые мы можем каким-то образом измерить, используя действительные числа. Для использования этого закона, нам необходим язык, достаточно богатый для формулировки не только самого закона, но и утверждений вида «сила f_{ab} равна $r_1 \pm r_2$ », «масса M_a равна $r_1 \pm r_2$ », «расстояние d равно $r_1 \pm r_2$ », где r_1, r_2 — любые

рациональные числа. (Совсем не обязательно, хотя и возможно, иметь для каждого отдельного действительного числа свое имя, но у нас обязательно должна быть возможность выразить любые оценки физических величин в рациональных числах.)

Однако ни один номиналист еще не предложил способа перевода произвольно взятых утверждений, имеющих форму «расстояние d равно $r_1 \pm r_2$ », на номиналистический язык. Более того, если мы не хотим постулировать существования реально бесконечного множества физических объектов, подобная «схема перевода» невозможна, о чем свидетельствует следующий простой аргумент: если множество индивидуальных объектов конечно, то в формализованном номиналистическом языке имеется только конечное множество парных неэквивалентных утверждений. Иными словами, имеется конечное множество утверждений S_1, S_2, \dots, S_n таких, что для произвольно взятого утверждения S : или $S \equiv S_1$, или $S \equiv S_2$, или ... $S \equiv S_n$; более того (для соответствующего i) $S \equiv S_i$ логически следует из утверждения «число индивидуальных объектов равно N »⁴. Однако, если наш

⁴ Приведем набросок доказательства этого утверждения: предположим, к примеру, что $N = 2$, и введем (времененно) символы « a » и « b » для двух, как мы полагаем, существующих индивидуальных объектов. Перепишем каждое предложение $(x) Px$ в виде конъюнкции $Pa \cdot Pb$, а каждое предложение $(\exists x) Px$ — в виде дизъюнкции $Pa \vee Pb$. Таким образом, каждое предложение S в нашем языке переведено в предложение без кванторов S' . Существует только конечное множество атомарных предложений (если предположить, что число исходных предикатов в языке конечно). Если число этих атомарных предложений равно n , то число функций истинности, которые можно для них записать, равно 2^{2^n} . Можно легко сформулировать 2^{2^n} предложений без кванторов, соответствующих этим 2^{2^n} функциям истинности; тогда любое предложение, построенное из данных n атомарных предложений с помощью истинностно-функциональных связей, будет логически эквивалентно одному из этих предложений $T_1, T_2, \dots, T_{2^{2^n}}$. Более того, если $S' \equiv T_i$ является теоремой пропозиционального исчисления, то легко показать, что $S \equiv (\exists a, b) (a \neq b \cdot T_i)$ истинно в любом двуэлементном универсуме, и, следовательно, из утверждения «число индивидов равно двум» (это можно выразить символически как $(\exists a, b) (a \neq b \cdot (x) (x = a \vee x = b))$) следует $S \equiv (\exists a, b) (a \neq b \cdot T_i)$. Итак, если мы принимаем, что $S_1 = \langle (\exists a, b) (a \neq b \cdot T_1) \rangle$, $S_2 = \langle (\exists a, b) (a \neq b \cdot T_2) \rangle$, ..., то (1) при числе индивидов, равном двум, каждое предложение S эквивалентно по своему истинностному значению одному из предложений $S_1, S_2, \dots, S_{2^{2^n}}$; и (2) предложение $S \equiv S_i$ (для соответствующего i) само *следует* из утверждения, что число индивидуальных объектов равно двум. Такое же рассуждение можно построить применительно к любому конечному числу индивидуальных объектов.

«язык физики» содержит имена для двух различных индивидуальных объектов, скажем, a и b , и мы можем формулировать утверждения «расстояние от a до b равно одному метру \pm один сантиметр», «расстояние от a до b равно двум метрам \pm один сантиметр» и т. д., то ясно, что мы должны иметь *бесконечный* ряд парных неэквивалентных утверждений. (Неэквивалентность не исчезает и при наличии посылки «число индивидов равно N »; она не вытекает логически из посылки, что любые два из вышеприведенных утверждений имеют одно и то же истинностное значение.) Таким образом, любой «перевод» «языка физики» в «номиналистический язык» с необходимостью нарушает логические отношения: для любого N можно найти два разных целых числа n , m таких, что ложная «теорема»:

«Если число индивидуальных объектов равно N , то «расстояние от a до b равно n метров \pm один сантиметр» \equiv «расстоянию от a до b равно m метров \pm один сантиметр», —

станет истинной теоремой логики, если мы принимаем указанную схему перевода. Итак, номиналистический язык, в принципе, не годится для физики.

Его непригодность станет еще очевидней, если мы посмотрим на проблему в менее формалистическом ракурсе. Понятие «расстояние в метрах» крайне сложно. Что включается в предположение о том, что такую физическую величину как расстояние можно каким-то образом соотносить с *действительными числами*?

Согласно одному объяснению (которое я считаю верным), физика обязывает нас признать существование таких сущностей, как «пространственные точки» (или пространственно-временные точки в релятивистской физике), хотя природа этих сущностей далеко не ясна. Физики часто утверждают, что пространственно-временные точки — это просто «события», хотя, очевидно, что это неверно. Карнап и Куайн предпочитают говорить о точках, как тройках действительных чисел (или тетрадах действительных чисел в случае пространственно-временных точек); однако это представляется весьма неестественным, поскольку в интуитивном смысле идентичность (identity) пространственной точки не зависит от конкретной системы координат. Я предпочитаю думать о них как о свойствах некоторых событий (или частиц, если в физике есть точки-частицы); но давайте на некоторое время будем считать их элементарными сущностями, которые можно идентифицировать только с помощью имени «точки». С любой точки зрения между точками имеет место физически зна-

чимое отношение $C(x, y, z)$, которое можно назвать отношением конгруэнтности; на обычном языке это отношение можно выразить так: интервал $\bar{x}y$ конгруэнтен интервалу $\bar{z}w$. (Я сказал «с любой точки зрения», потому что существуют серьезные разногласия между философами: одни полагают, что это отношение можно определить операционально, а другие, к коим принадлежу и я, считают, что все так называемые операциональные определения далеко неточны и что рассматриваемое отношение следует признать исходным в физической теории). Возьмем две точки (скажем, конечные точки эталона метра в Париже в конкретный момент времени) и назовем их a_1 и a_2 . Примем, что расстояние от a_1 до a_2 , по определению, равно единице. Тогда «расстояние» между любыми двумя точками x и y можно определить как некую числовую величину следующим образом:

Согласно определению, «расстояние от x до y равно r » означает, что $f(x, y) = r$, где f — любая функция, удовлетворяющая следующим пяти условиям:

(1) $f(w, v)$ определена (и имеет в качестве значения неотрицательные действительные числа) для любых точек w, v .

(2) $f(w, v) = 0$, если и только если w та же самая точка, что и v .

(3) $f(w, v) = f(w', v')$, если и только если выполняется отношение $C(w, v, w', v')$ (т. е. если и только если интервал wv конгруэнтен интервалу $w'v'$).

(4) Если w, v, u — коллинеарные точки, и v находится между w и u , то $f(w, u) = f(w, v) + f(v, u)$. (Понятия «коллинеарный» и «между» можно или определить стандартным способом в терминах отношения C , или взять как исходные понятия физической геометрии.)

(5) $f(a_1, a_2) = 1$.

Можно показать, что существует только одна функция, удовлетворяющая условиям (1)—(5)⁵. Стало быть, содержание вышеприведенного определения можно выразить так: расстояние — это значение единственной функции, удовлетворяющей условиям (1)—(5).

⁵ Строго говоря, это истинно только, если f является непрерывной функцией, отображающей пространственные точки на действительные числа. Однако это требование непрерывности можно соблюсти, даже если мы не предполагаем, что у нас уже есть метрика, применимая к пространственным точкам. Я опустил этот момент в настоящей статье в целях упрощения изложения.

Назовем то, что было описано выше, «нумеризацией» (numericalization) ⁶ физической величины расстояния. В контексте нашего рассмотрения главным является следующее: даже если мы рассматриваем «точки» как отдельные индивидуальные объекты, а отношение « $S(x, y, z, w)$ » — как исходное, мы все равно не сможем объяснить нумеризацию расстояния без помощи квантификации по функциям. (Конечно, мы можем уйти от проблемы, отождествив точки с тройками действительных чисел и воспользовавшись теоремой Пифагора, чтобы сформулировать определение расстояния, но тогда мы или должны были бы проанализировать отношение «объект O находится в точке P », или должны счесть нумеризацию совершенно загадочной и необъяснимой процедурой.)

Короче говоря, даже утверждение вида «расстояние от a до b равно $r_1 \pm r_2$ », где r_1 и r_2 — переменные, определенные на множестве рациональных чисел, нельзя объяснить без использования понятия функции, отображающей точки на действительные числа или, по крайней мере, на рациональные числа. Если для любых констант r_1 и r_2 можно сформулировать эквивалентное утверждение с квантификацией только по «точкам», то, чтобы объяснить значение предиката, когда он применяется к переменным r_1 и r_2 , необходимы понятия функции или множества. И даже обычное решение, как мы только что видели, предполагает функции, отображающие точки на действительные числа.

Одному и тому же человеку нетрудно в одном контексте придерживаться номиналистических убеждений, а в другом — говорить о расстоянии как о чем-то определенном (и обладающем числовым значением) для любых произвольно взятых точек x и y . Однако, как мы только что видели, это непоследовательно. Если нумеризация физических величин имеет какой-то смысл, то мы должны принять такие понятия, как функция и действительное число; но именно эти понятия номиналисты и отвергают. Если этим понятиям ничего не соответствует, то что же тогда говорит закон гравитации? Этот закон не будет иметь никакого смысла, если мы не сможем объяснить

⁶ В любом тексте по философии науки вы найдете термин «измерение», а не «нумеризация». Я выбрал этот варварский термин, ибо хотел подчеркнуть, что проблема состоит не в том, чтобы *измерить* что-либо, а в том, чтобы *определить* нечто, а именно — определить соответствие между парами точек и числами. Термин «измерение» — это наследие времен операционализма, когда предполагалось, что измерение предшествует определению, а не наоборот.

переменные, значениями которых являются произвольно взятые расстояния (а также силы и массы).

VI. ПРЕДИКАТИВНАЯ ИЛИ НЕПРЕДИКАТИВНАЯ КОНЦЕПЦИИ «МНОЖЕСТВА»

Множество $\{x, y\}$, содержащее только два элемента x, y называется неупорядоченной парой x и y . На основе неупорядоченных пар можно различными способами определить упорядоченные пары. Наиболее естественный, хотя и не самым привычный, способ состоит в следующем: Возьмем в качестве «маркеров» два объекта a и b . Определим упорядоченную пару x и y как множество $\{\{x, a\}, \{y, b\}\}$, т. е. как неупорядоченную пару, элементами которой являются неупорядоченные пары $\{x, a\}$ и $\{y, b\}$. Введем для этой упорядоченной пары обозначение $\langle x, y \rangle$, т. е. по определению $\langle x, y \rangle$ есть $\{\{x, a\}, \{y, b\}\}$. Теперь легко понять, что для любых x, y, u, v :

$$\langle x, y \rangle = \langle u, v \rangle,$$

если и только если $x = u$ и $y = v$. Таким образом, две «упорядоченные пары» тождественны только в том случае, если они состоят из одних и тех же элементов и эти элементы одинаково упорядочены — только это и требуется для определения упорядоченной пары.

В математике двуместное отношение — просто множество упорядоченных пар. Поскольку «упорядоченная пара» была только что определена на основе «неупорядоченной пары», а «неупорядоченная пара» — это просто множество, то отсюда следует, что «отношение» можно определить на основе одного исходного понятия *множества*. Пусть R отношение такое, что для всех u, v, y

$$\text{если } \langle u, v \rangle \in R \text{ и } \langle u, y \rangle \in R, \text{ то } v = y,$$

тогда это отношение R называется «функцией». Поскольку функция была только что определена на основе понятия отношения (и понятия « $=$ », которое мы считаем принадлежащим к элементарной логике), то отсюда следует, что и функция определяется на основе понятия *множества*.

Хорошо известно, что натуральные числа 0, 1, 2, 3, ... можно разными способами определить на основе понятия *множества*. Например, можно определить 0 как пустое множество, 1 — как множеством $\{0\}$, 2 — как $\{0, 1\}$, 3 — как $\{0, 1, 2\}$ и т. д. Кроме того, и все элементарные операции «сложить», «умножить» и т. п., можно опре-

делить на основе понятия *множества*. Рациональные числа естественным образом определяются как упорядоченные пары натуральных чисел, не имеющих общего делителя (при этом второй член упорядоченной пары не равен нулю); действительные числа можно, например, определить как ряды рациональных чисел, где «ряд» — функция, областью определения которой являются натуральные числа. Таким образом, все «объекты» чистой математики можно построить, опираясь на понятие *множества*; и именно так предпочитают делать все современные математики.

Таким образом, если в предыдущем разделе мы говорили, что физика не может обойтись без ссылок на функции и действительные числа, то теперь мы можем просто сказать, что физике необходимо лишь такое понятие как *множество*, поскольку понятия числа и функции можно построить на основе понятия множества. В настоящем разделе мы дадим беглый обзор понятия множества.

Наиболее известная трудность в отношении понятия множества такова. Предположим, что:

(1) Множества — это самостоятельные сущности (т. е. сущности, по которым мы можем квантифицировать ⁷);

(2) Если \emptyset — любое точно определенное условие, то существует множество всех сущностей, удовлетворяющих условию \emptyset .

Тогда (допуская также, что условие « $\sim x \in x$ » точно определено) мы выводим, что существует множество всех множеств x таких, что x не принадлежит к x . Пусть y — такое множество, тогда:

(3) $(x) (x \in y \equiv \sim x \in x)$.

Затем, подставляя y вместо x , получаем

(4) $y \in y \equiv \sim y \in y$,

а это — противоречие!

Очевидно, что одно из наших допущений было ложным. Какое же? Мы могли бы сказать, что « $\sim x \in x$ » не является точно определенным условием. Однако, если $x \in y$ точно определенное отношение для любых произвольно выбранных множеств x и y , то представляется, что и $x \in x$, и $\sim x \in x$ должны быть точно определены (в том смысле, что они имеют определенное истинностное значение) для любого множества x . Отказ считать, что $x \in y$ — точно опреде-

⁷ «Квантифицировать по» множествам значит использовать такие выражения, как «для каждого множества x » и «существует множество x такое, что».

ленное отношение или что множества — это самостоятельные сущности, означал бы отказ от самой теории множества. В таком случае единственная альтернатива — это отказаться от (или, по крайней мере, ограничить) условие (2), которое сильно расходится с нашими интуициями.

Одним из способов разрешить эту трудность является так называемая теория типов. Согласно этой теории, « $x \in y$ » точно определено, если и только если x и y относятся к соответствующим типам; при этом считается, что индивидуальные объекты принадлежат к типу 0, множества индивидуальных объектов — к типу 1, множества множеств индивидуальных объектов к типу 2 и т. д. Согласно этой теории, выражение « $\sim x \in x$ » нельзя даже считать грамматически правильным, поскольку ни об одном множестве мы не можем сказать, что оно является или не является собственным элементом. Можно говорить о том, принадлежит ли множество к любому множеству более высокого типа, но нельзя говорить о том, принадлежит ли множество к самому себе (или к любому другому множеству, не относящемуся к более высокому типу).

Пусть R — некоторое отношение между индивидуальными объектами. Назовем R -цепью такое множество α , что для любого x , если $x \in \alpha$, то существует хотя бы один y такой, что Rxy и $y \in \alpha$. Допустим, имеется некоторая R -цепь, содержащая индивидуальный объект U . Тогда запишем:

$$(5) (\exists \alpha) (\alpha \text{ есть } R\text{-цепь. } U \in \alpha),$$

где « α есть R -цепь» является краткой записью для « $(x) (x \in \alpha \supset (\exists y) (y \in \alpha \cdot Rxy))$ ».

Пусть β — множество всех U таких, что некоторая R -цепь содержит U . Это множество совершенно законно с точки зрения теории типов, и его принимает большинство математиков. Однако отдельные математики и философы возражают против идеи такого множества. По их мнению, определение множества β как множества всех U таких, что существует R -цепь, содержащая U , «порочно», поскольку «совокупность (totality), в терминах которой определяется β », т. е. совокупность α , содержащая все R -цепи могла бы содержать и само множество β . Как правило, эти математики и философы запрещают определять множество в терминах «совокупности», если только мы не удостоверились, что эта совокупность не способна содержать это множество или любое другое, определенное с помощью

этого первого множества. Конечно, в этом много неясного. Однако представляет интерес мотив, обусловивший такое решение.

Предположим, я вообще не понимаю слово «множество» и, по существу, использую только некоторый номиналистический язык N . Однажды я прихожу к выводу, что знаю два понятия, которые не являются номиналистическими или, во всяком случае, имеют спорный номиналистический статус, а именно: понятия «формула» и «истина». На основе этих понятий я могу ввести очень слабое понятие множества — определяя множества как формулы моего номиналистического языка, содержащие только одну свободную переменную x ; например, я определяю множество красных вещей как формулу «Красное (x)». Понятие «быть элементом» я объясняю следующим образом: если y — индивидуальный объект, а α — «множество» (т. е. формула с одной свободной переменной « x »), то « $y \in \alpha$ » означает, что α истинно относительно y ; при этом формула $\emptyset(x)$ считается истинной относительно индивидуального объекта y только в том случае, если она истинна, когда x интерпретируется как имя для y . Итак, если α — формула «Красное (x)», то мы имеем:

- $y \in \alpha$, если и только если α истинно относительно y ,
- т. е. если и только если «Красное (x)» истинно относительно y ,
- т. е. если и только если y есть красное.

Поэтому «Красное (x)» оказывается «множеством всех красных вещей», как это и должно быть.

Это понятие множества я называю «слабым», поскольку оно заставляет говорить обо *всех* множествах индивидуальных объектов, а тем более — о множествах, имеющих тип выше, чем 1; конечно, можно говорить обо всех формулах, но это означает лишь — говорить обо всех множествах индивидуальных объектов, определяемых в языке N . Если в N будут введены новые исходные выражения, то в целом совокупность множеств, как они были только что определены, увеличится. Однако, можно повторить вышеописанную процедуру. Пусть N' — язык, получаемый из N , если разрешить квантификацию по всем множествам индивидуальных объектов, определяемых в N , N'' — язык, получаемый из N' , если разрешить квантификацию по всем множествам индивидуальных объектов, определяемых в N' , и т. д. Тогда все эти множества индивидуальных объектов, определяемых в N , N' , N'' , ... представляют собой примеры «предикативных» множеств: каждое из этих множеств предполагает «совокупность», которая определена «ранее» (начиная с совокупности индивидуальных

объектов), но которая не предполагает самого этого множества. (Можно также ввести предикативные множества более высокого типа на основе понятия формулы о формулах, но мы не будем этого делать.) Для нас здесь важно подчеркнуть следующее: предикативное понятие множества можно разъяснить применительно к любому языку из серии N, N', N'', \dots , используя понятие квантификации, применимое только к тем множествам, которые были ранее определены в указанной серии, и в целом эту терминологию — «множества, определяемые в N », «множества, определяемые в N' » и т. д. — можно считать, если угодно, просто *façon de parler*⁸, который можно объяснить на основе понятий формулы и истины.

Если в противовес сказанному ранее мы продолжаем говорить не только обо всех множествах, определяемых в некотором языке серии N, N', N'', \dots , но и обо всех множествах индивидуальных объектов как о точно определенной совокупности, то считается, что мы используем непредикативное понятие множества.

VII. НАСКОЛЬКО ТЕОРИЯ МНОЖЕСТВ ДЕЙСТВИТЕЛЬНО НЕОБХОДИМА НАУКЕ?

Ранее мы утверждали, что понятие множества (или некоторое эквивалентное понятие, например, понятие функции) необходимо науке. Однако, теперь мы должны спросить: какое понятие множества необходимо науке — «строгое» (непредикативное) или только «слабое» (предикативное)? Если мы действительно хотим разобраться с номинализмом и реализмом, то нам следует допустить в качестве альтернатив не только (а) номинализм, но и (б) принятие общего понятия «все множества» (или даже «все множества индивидуальных объектов»). Если у нас есть склонность к номинализму, то мы стараемся сделать свои неноминалистические обязательства как можно более слабыми, а потому возможность ограничить эти обязательства двумя понятиями истины и формулы представляется весьма желательной. Некоторые номиналисты считают истину понятием, на которое они в любом случае имеют право; а формулы (трактуемые как типы формул, представленные реальными записями), если и являются «абстрактными» сущностями, то все-таки относительно ясными.

В случае чистой математики можно было бы ответить следующее: определенную часть математики можно разрабатывать, исполь-

⁸ Речевой оборот (франц.) — Прим. перев.

зую только предикативную теорию множеств, при условии, что допускаются предикативные множества других, не только физических, объектов. Например, если мы рассматриваем формулы языка N в качестве индивидуальных объектов для некоторого другого языка M , а затем строим серию языков M, M', M'', \dots по вышеописанной схеме, то мы можем, по крайней мере, сформулировать арифметику рациональных чисел и элементарную теорию функций для рациональных чисел. (Однако, «для начала» нам необходима некоторая бесконечная область индивидуальных объектов; вот почему мы вынуждены рассматривать абстрактные объекты, например, формулы, в качестве индивидуальных объектов, если не хотим постулировать существование реально бесконечной совокупности физических объектов.) К сожалению, этим способом нельзя получить никакой удовлетворительной теории действительных чисел и теории функций для действительных чисел; по этой причине большинство математиков отвергают предикативную теорию.

Возвратимся к логике, к понятию «корректности». Выше мы уже отмечали, что понятие корректности, а именно — «истинности всех результатов подстановок» (скажем, в M), можно, по сути дела, определить в терминах предикативной теории множества (т. е. на основе понятий истины и квантификации по формулам). Мы также видели, что для более удовлетворительного понятия корректности требуется использование выражений «все множества», т. е. понятия непредикативной теории множества.

И, наконец, возвращаясь к физике, мы видим следующее. На первый взгляд, закон всемирного тяготения (здесь предполагается, что это единственный закон физики) требует квантификации по *действительным* числам. Однако, этот закон эквивалентен утверждению, что для каждого рационального числа ε и любых рациональных чисел m_1, m_2, d существует рациональное число δ такое, что

$$\text{если } M_a = m_1 \pm \delta, M_b = m_2 \pm \delta, d = d_1 \pm \delta,$$

то

$$F = \frac{g m_1 m_2}{d_1} \pm \varepsilon$$

а это утверждение предполагает квантификацию только по рациональным числам. Хотя (!) гравитационная постоянная g может и не быть рациональным числом, но я не буду касаться здесь этой про-

блемы.) Таким образом, язык, допускающий квантификацию только по рациональным числам и позволяющий измерять расстояния, массы, силы и т. п. только с рациональной аппроксимацией («масса a равна $m_1 \pm \delta$ »), в принципе, является достаточно строгим, по крайней мере, для формулировки закона всемирного тяготения.

Рациональные числа можно легко определить, используя только предикативную теорию множеств. Этой теории множеств также достаточно для определения «кардинального числа S », где S — любое определимое конечное множество физических объектов. Осуществить «нумеризацию» таких физических величин, как расстояние, сила, масса, используя рациональные аппроксимации и предикативные множества, достаточно сложно, но вполне возможно. Итак, при использовании только предикативной теории множеств физика оказывается возможным (но сложным и неудобным) занятием.

Подведем итоги. Теоретико-множественные «нужды» физики удивительно совпадают с теоретико-множественными «нуждами» чистой логики. Обеим этим дисциплинам для их функционирования нужна *какая-то* теория множества. Обе дисциплины могут «существовать», но с трудом, имея в качестве «скудной диеты» только предикативные множества, но их существование становится гораздо привлекательней при «обильной диете» в виде непредикативных множеств. Поскольку необходимость квантификации по множествам является аргументом в пользу их существования (в следующем разделе мы обсудим, почему это так), поэтому мы можем сказать, что это строгий аргумент в пользу существования, по крайней мере, предикативных множеств и достаточно, хотя и *не столь* строгий, аргумент в пользу существования непредикативных множеств. Однако, обращаясь к более высоким уровням теории множества — множествам множеств множеств множеств, мы сталкиваемся с понятиями, в которых сегодня нуждается только чистая математика. Таким образом, развиваемый в настоящем разделе аргумент в пользу «реализма» имеет ограниченный характер: по крайней мере, множества предметов, действительные числа и функции, отображающие различного рода предметы на действительные числа, следует признать составляющими на сегодня необходимую (или практически необходимую) структуру как физической науки, так и логики, и в этом качестве мы обязаны признавать их существование. Но множества очень высокого типа или очень большой мощности (например, большей мощности континуума) следует сегодня изучать с оговоркой «если». Когда-нибудь они могут стать такими же необходимыми для формулировки физических законов, какими сегодня являются, скажем, ра-

циональные числа, и тогда сомнения в их «существовании» будут столь же тщетны, как и крайний номинализм сегодня. Но на сегодняшний день мы должны видеть в них то, чем они являются, а именно — умозрительными и дерзкими сооружениями над основным математическим аппаратом науки.

VIII. НЕОБХОДИМОСТЬ КАК АРГУМЕНТ В ПОЛЬЗУ РЕАЛИЗМА

До сих пор я строил свой аргумент в пользу реализма примерно по такой схеме: квантификация по математическим сущностям необходима как физической, так и формальной наукам, следовательно мы должны принять квантификацию, а это обязывает нас признать существование упомянутых математических сущностей. Конечно, этот аргумент идет от Куайна, который в течение многих лет подчеркивал необходимость квантификации по математическим сущностям и усматривал интеллектуальную нечестность в отрицании существования того, что постоянно предполагается в рассуждениях. Однако указанные аргументы поднимают ряд вопросов; краткому обсуждению некоторых из них я и посвящаю этот раздел.

Один из этих вопросов касается, например, осмысленности таких предложений, как «существуют числа», «существуют множества», «существуют функции, отображающие пространственно-временные точки на действительные числа» и т. п. Если это не подлинные утверждения, а, так сказать, псевдо-утверждения, то никакими аргументами их нельзя будет обосновать — *a fortiori* ссылками на «необходимость» указанных сущностей для науки.

Но какие есть основания считать предложения «существуют числа», «существуют множества» и т. п. неосмысленными? Можно предположить, что с этими «утверждениями» должно быть *что-то* не так, поскольку мы сталкиваемся с ними только в философии. Но, в этом способе рассуждения есть что-то крайне сомнительное, несмотря на всю его популярность в наши дни. Одно дело *показать*, что выражения, в которых формулируется определенная философская проблема с лингвистической точки зрения девиантны (*deviant*). Действительно, если нельзя найти *никакого* способа сформулировать предлагаемую «проблему», который *не* искажал бы языка, то оправданно заподозрить «проблему» в ее совершенной неясности, но даже и в этом случае этого нельзя утверждать с достоверностью, потому что девиантные в лингвистическом плане выражения *необязательно* буквально лишены смысла (*unintelligible*). Однако даже если ключевые термины предполагаемой философской проблемы в линг-

лингвистическом плане девиантны (или, выражаясь неформальным языком, «странны», «сомнительны» и т. п.), то это отнюдь не доказывает неподлинность этой проблемы, тем более, если эта «девиантность» (или «странность», «подозрительность» и т. п.) была установлена на основе того сомнительного принципа, что термины и утверждения, встречающиеся только в философии, *ipso facto* девиантны. Ибо трудность (а на самом деле, больше, чем просто трудность) состоит в том, что для этого ошеломляющего заявления нет *лингвистических* оснований. Каждая дисциплина имеет свои специфические термины и утверждения, и нет причин отказывать в этом философии. Если, например, утверждение «существуют материальные объекты» не встречается за пределами философии, так это потому, что только философов интересует, что дает нам основания верить в это очевидное утверждение и только философы обладают терпением и профессиональными навыками, необходимыми для анализа обоснований, который оказывается столь трудной задачей; какую другую науку интересует обоснование и оправдание как таковые? Хотя часто можно услышать, что философские высказывания по самой своей сути лингвистически (логически или «концептуально») неясны, однако нет ни малейшего *лингвистического* свидетельства о том, что такие предложения, как «существуют числа», «существуют множества» и «существуют материальные объекты», девиантны в лингвистическом плане, т. е. нарушают определенные нормы естественного языка, в отношении которых можно с помощью соответствующих научных процедур установить, что они *являются* нормами естественного языка.

Иными словами, если бы мы могли честно *показать*, что выражения, специфические для философского дискурса в лингвистическом плане неправильны, это было бы важным и ошеломляющим результатом. Однако этот результат не представляет никакого интереса, если его «обосновывают» лишь тем, что определенные выражения, специфические для философии, *должны* быть в чем-то неправильными, *потому что* они являются специфическими для философии и *потому что* выражения, встречающиеся только в философском дискурсе, «странны». Это доказательство содержит явный круг: выдвигается принцип *P* (гласящий, что есть нечто неправильное в выражениях, встречающихся только в философском дискурсе), в его подтверждение предлагается множество примеров (т. е. философских утверждений и вопросов, которые якобы «странны», «сомнительны» и т. д.); но, оказывается, что эти примеры подтверждают только в том случае, если принят принцип *P*. Я не отрицаю, что исторически многие философские утверждения и аргументы содержали

«сомнительные», с точки зрения любого критерия, выражения (а в некоторых случаях существенным образом зависели от них). Я утверждаю лишь, что с лингвистической точки зрения нет ничего «сомнительного» ни в общих вопросах существования *per se* («Существуют ли числа?», «Существуют ли материальные объекты?»), ни в общих вопросах оправдания или обоснования («Какие есть основания верить в то, что материальные объекты существуют?»). (Например, вторые вопросы были отвергнуты на основе только что описанного рассуждения, содержащего «порочный круг», в книге Дж. Л. Остина «Чувство и сенсibiliа» («*Sense and Sensibilia*»), и я уверен, что многие философы сходным образом отвергли бы и первые вопросы.)

До сих пор я пытался доказать, что нет иной причины относить утверждения типа «существуют числа» и «существуют множества» к девиантным в лингвистическом плане, кроме как на основе философского принципа, который, видимо, только вводит в заблуждение. Кроме того, есть простой способ полностью обойти этот вопрос. Даже если некоторые философы отвергают предложение «существуют числа», как не относящееся к нормальному языку, предложение «существуют числа, имеющие свойства —» считается правильным (и даже истинным) при многих значениях «—». Например, предложение «существуют простые числа больше 10^{10} », безусловно, правильно и истинно. Тогда, если даже предложение «существуют числа» *simpliciter* не относится к нормальному языку, мы всегда можем ввести его в язык просто как новый речевой оборот, оговорив следующее условие: предложение «существуют числа» истинно, если и только если имеется свойство «—» такое, что предложение «существуют числа, имеющие свойство “—”» истинно.

Это равносильно следующему: если предложение

$$(1) (\exists x) (x \text{ есть число} \cdot x \text{ есть простое число} \cdot x > 10^{10})$$

(т. е. предложение, имеющее такое символическое выражение) относится к нормальному языку, тогда предложение

$$(2) (\exists x) (x \text{ есть число})$$

(означающее «существуют числа») не относится к нормальному языку, то обыденный язык не является «дедуктивно замкнутым»: поскольку предложение (2) выводимо из предложения (1) средствами обычной логики (с помощью теоремы « $(\exists x) (Fx \cdot Gx \cdot Hx) \supset (\exists x) Fx$ »). Но если обыденный язык в этом аспекте дедуктивно не замкнут, то мы можем исправить это, введя в него предложение (2); более того, это

можно сделать только таким способом. Поэтому мы можем считать предложение (2) правильным выражением языка.

Итак, мы опровергли ту точку зрения, что предложения «существуют числа», «существуют множества» и т. п., девиантны в лингвистическом плане, не обладают истинностным значением и т. п.

Другая причина, почему некоторые философы могли бы не принимать аргументы, ссылающиеся на необходимость абстрактных сущностей для науки, такова: эти философы утверждают, что логические и математические истины представляют собой *конвенциональные истины*. Если, например, предложения «существуют числа» и «существуют множества» — это конвенциональные истины, то соображения необходимости *не имеют никакого значения*.

Однако эта «конвенционалистская» позиция терпит крах, как только конвенционалиста просят уточнить свою позицию. *Каким именно образом с помощью понятия конвенции можно определить понятие истины применительно к предложениям, содержащим кванторы по абстрактным сущностям? Даже если предположить, что некоторые математические предложения действительно «конвенциональные истины» в том смысле, что их истинность непосредственно устанавливается конвенцией, и если допустить, что эти предложения можно перечислить, все равно для объяснения других математических истин, которые в любом случае не являются непосредственно конвенциональными, а требуют доказательства, конвенционалисту потребуется понятие импликации. Однако для определения понятия импликации (как корректного условного высказывания) необходима, как мы видели, теория множеств. Стало быть, конвенционализм, даже если он верен, предполагает, помимо понятия конвенции, квантификацию по абстрактным сущностям; в итоге оказывается, что для объяснения математической истины нужна конвенция плюс математика, а это довольно тривиальное объяснение (я не касаюсь важного вопроса о том, насколько велик элемент конвенциональности в математике). Более того, если конвенционалист не соблюдает осторожности, его теория математической истины может легко прийти в противоречие с результатами самой математики, в частности с теоремой Гёделя. Однако, дальнейшее обсуждение этого вопроса увело бы нас в сторону, и поэтому сейчас я просто отброшу конвенционализм на том основании, что никому еще не удалось дать такую формулировку этой точки зрения, которая была бы точной и не потеряла бы сразу же крах.*

В качестве третьей причины для отказа от предложенных нами аргументов можно было бы выдвинуть следующую: как утверждали

на рубеже веков ряд философов, разнообразные сущности, предполагаемые в научных и иных рассуждениях, — а по мнению некоторых этих философов, и сами материальные объекты, — представляют собой просто «полезные фикции», или во всяком случае, полагали эти философы, мы не можем знать, что эти сущности есть нечто *большее*, чем «полезные фикции» (поэтому мы и можем говорить о них как о фикциях). Сегодня этой философии «фикций», видимо, никто уже не придерживается, но ее следует здесь затронуть хотя бы потому, что она напрямую отрицает доказательную силу рассматриваемых здесь аргументов. Сторонник фикционализма говорит, по сути дела, следующее: «Да, некоторые понятия — материальный объект, число, множество и т. д. — необходимы, но это *не означает*, что сущности, соответствующие этим понятиям, действительно существуют. Это доказывает только, что эти «сущности» — *полезные фикции*».

Хотя фикционализм и был опровергнут современными философами науки и эпистемологами, но, думается, отчасти — на ошибочных основаниях. Представители фикционализма считали как логически возможное следующее: если в реальности и не может существовать электронов (и тому подобного), то наш опыт может быть таким, как если бы они реально существовали. Согласно популярному с конца 1920-х годов верификационизму, это утверждение *бессмысленно*: если p — высказывание, которое логически невозможно верифицировать, то p не может выражать логическую возможность. Однако в этом вопросе представители фикционализма были совершенно правы, а верификационисты — ошибались: было бы абсурдно, безумно, нелепо и совершенно нерационально верить, например, в то, что все мы бестелесные духи, управляемые мыслью некоего могущественного разума, главная цель которого — внушить нам ложную идею, будто существует материальный мир; однако это предположение не является *логически невозможным*. Я не занимаюсь здесь анализом верификационизма, но мимоходом замечу, что все аргументы верификационизма оказались ошибочными. Конечно, его главный аргумент состоял в том, что предложение «существуют материальные объекты» *означает* только то, что при определенных обстоятельствах мы имеем определенный опыт; однако все попытки выполнить программу действительно полной и адекватной редукции предложений о материальных объектах к предложениям о «чувственных данных» потерпели крах, и сегодня почти с полной уверенностью можно сказать, что такая редукция невозможна. Если дана достаточно богатая теория T , содержащая как предложения о «чувственных данных», так и «предложения вещного языка» (допустим, снисходи-

тельности ради, что язык «чувственных данных» действительно может быть построен), то можно было бы сказать, какие предложения о «чувственных данных» логически выводимы из T ; но это не означает, что предложения вещного языка, сформулированные в T (а в еще меньшей степени, сформулированные вне какой-либо теории), должны быть эквивалентны — каждое по отдельности — предложениям о чувственных данных в любом приемлемом смысле «эквивалентности». Другим верификационистским аргументом является аргумент в форме риторического вопроса: «Что еще может означать утверждение о том, что материальные объекты существуют, кроме того, что при таких-то и таких-то условиях мы обычно имеем такой-то и такой-то опыт?» Однако аргумент в форме риторического вопроса предполагает успешную феноменалистическую редукцию. Если у вас есть перевод S' предложения вещного языка S на феноменалистический язык, то вы вполне можете спросить: «Что еще, помимо S' , означает S ?», но вы не можете задавать этот риторический вопрос, если у вас нет перевода S' . Согласно еще одному верификационистскому аргументу: «псевдо-гипотезы, вроде гипотезы о демоне, обладают только *значением-образом* (picture meaning)». Во-первых, вызывает возражения форма этого аргумента (который, формулируя философский вопрос, дает лишь психологическое объяснение склонности наших оппонентов к ошибкам), а, во-вторых, он просто ложен. «Гипотеза о демоне» — это не просто звук, рождающий «в голове различные образы»; это грамматически правильное предложение, которое мы можем легко переводить на другие языки и можем формулировать по-разному; мы можем выводить из него другие предложения и можем сказать, из каких предложений следует оно само; мы можем установить, уместно ли оно в том или ином лингвистическом контексте, правилен или девиантен в лингвистическом плане содержащий его дискурс. Верификационисты возразили бы: «Отсюда не следует, что оно имеет *значение*». Однако, здесь они не правы, ибо именно это и есть значение: иметь значение значит допускать определенные рекурсивные преобразования и подчиняться определенным закономерностям; возможно, сегодня мы и не знаем большего в этом вопросе, но мы знаем достаточно и понимаем, что предложенное верификационистами — это не анализ значения, а убедительное переопределение. Однако самым слабым верификационистским аргументом был следующий: «Если вы *действительно* принимаете гипотезу о демоне как логическую возможность, то вы неизбежно скатитесь в скептицизм, ибо никогда не сможете обосновать ложность этой гипотезы». Если нам заявляют, что верифика-

ционизм и только верификационизм сможет уберечь нас всех от болота скептицизма, то мы можем ответить следующее (возможно, кому-то будет интересно это услышать): «Если гипотеза о демоне сформулирована таким образом, что имеет те же самые проверяемые следствия, что и более правдоподобная система гипотез, которую мы принимаем (или имеет те же самые проверяемые следствия, что и любая система гипотез, которую любой рациональный человек счел бы более правдоподобной), то она не является логически ложной, но логически невозможно, чтобы когда-нибудь было рационально верить в нее. Ибо рациональность предполагает, что, если две гипотезы H_1 и H_2 дают одни и те же проверяемые предсказания (всегда или в настоящий момент) и гипотеза H_1 *a priori* гораздо более правдоподобна, чем H_2 , то следует отдать предпочтение гипотезе H_1 , а не H_2 . Если же была принята гипотеза H_1 , а все другие гипотезы, *a priori* более правдоподобные, чем H_1 , имели ошибочные следствия, то мы не должны отбрасывать гипотезу H_1 только потому, что кто-то указывает нам на логическую возможность ее ложности. (Приблизительно так формулируется «правило 4» в «Началах» Ньютона.)

Но возникает вопрос: «А есть ли на самом деле такая вещь, как правдоподобность *a priori*?». Нетрудно показать, что любая индуктивная логика явно или неявно предполагает *априорное* упорядочение гипотез с учетом их «простоты», вида содержащихся в них предикатов, формы предлагаемых ими законов или с учетом любых других факторов. Если мы вообще отказываемся *априорно решать*, какие гипотезы более или менее правдоподобны, то мы исключаем для себя возможность каких бы то ни было индуктивных экстраполяций на основе прошлого опыта; ибо в любой момент времени имеется бесконечное множество взаимно несовместимых гипотез, каждая из которых согласуется с некоторым конечным набором данных; когда мы объявляем какую-то гипотезу «подтвержденной», то мы делаем это не потому, что мы исключили *все* другие гипотезы, а потому, что сочли эти остальные гипотезы неправдоподобными, хотя они могут соответствовать фактам и даже предсказывать их, т. е. в какой-то момент следует отбрасывать гипотезы на *априорных* основаниях с тем, чтобы можно было принять какую-то одну гипотезу. Скептик возразит нам: «Как вы можете знать, что гипотеза о демоне менее правдоподобна, чем обычная гипотеза?». В ответ укажем, что упорядочить гипотезы по правдоподобности — это не значит высказать суждение об эмпирическом факте или сформулировать теорему дедуктивной логики; это значит выбрать методологическую позицию. Только выбрав такую методологическую позицию, можно сказать,

является ли гипотеза о демоне «безумной» или нет; я описал выбранную мной позицию (и как выразитель этой позиции добавлю, что этой позиции явно или неявно придерживается любой рациональный человек). В итоге, мы можем «исключить» гипотезу о демоне и нам не нужно для этого жонглировать понятиями логической невозможности или бессмысленности; мы должны только признать, что выбрали позицию, согласно которой эта гипотеза *a priori* менее вероятна, чем наша обычная гипотеза, а затем отметить своеобразный факт: логически истинно (в силу способа построения самой гипотезы о демоне) то, что если гипотеза о демоне истинна, то вера в нее не может быть рациональной (конечно, здесь предполагается следующая максима рациональности: не верить в гипотезу H_1 , если все явления, объясняемые с помощью гипотезы H_1 , объясняет и гипотеза H_2 , но H_2 более правдоподобна, чем H_1). Но если логически истинно (с учетом вышеназванной максимы рациональности), что иррационально когда бы то ни было верить в гипотезу о демоне, то этого достаточно; а раз мы можем обосновать, почему эту гипотезу следует отбросить, то нет необходимости доказывать, что она даже не представляет собой логической возможности.

Другой популярный способ опровержения фикционализма уходит своими корнями не в верификационизм, а в инструментализм. Здесь мы сталкиваемся со следующим ходом рассуждений: вопрос о том, являются ли утверждения «истинными», нельзя отделить от вопроса о том, рационально ли принять эти утверждения (пока с этим вполне можно согласиться), ибо утверждение «*p* истинно» считается рационально приемлемым только в том случае, если рационально приемлемо *p*. Однако конечная цель всей нашей «концептуальной системы» — это просто предсказание и упорядочение опыта (и, возможно, «простота», что бы она ни означала). Представитель фикционализма признает, что концептуальная система, предполагающая материальные объекты (или какие бы то ни было объекты), позволяет делать успешные предсказания (или успешные в той мере, в какой нам удалось определить их время) и является достаточно простой, чтобы мы могли определить время своих предсказаний. Но именно эти факторы и определяют рациональную приемлемость, поэтому рационально принять нашу концептуальную систему и назвать все высказывания, которые в нее входят, «истинными» (или «такими же истинными, как и все остальное», по удачному выражению Энтони Квинтона, ибо мы всегда сохраняем за собой право изменить свое мнение).

Несомненно, в этом возражении против фикционализма есть очень верная мысль. Стоит напомнить представителю фикционализма

ту элементарную истину, что мы не можем разделить основания, по которым рационально принять высказывание *p*, и основания, по которым рационально принять утверждение «*p* истинно». Сам я предпочитаю не говорить о простоте, поскольку простота в любом допускающем измерение смысле (например, оцениваемая по длине выражений, числу логических связок или числу аргументных мест в предикатах) — это только *один*, причем не самый важный, фактор, влияющий на суждения ученых и рациональных людей об относительной правдоподобности. Однако не это главное; мы можем просто признать, что инструменталист использует слово «простота» для обозначения сложного предмета, зависящего от многих факторов, и неважно, что некоторые коннотации вводят в заблуждение. Представитель фикционализма принимает, что предсказательная сила и «простота» (т. е. общая правдоподобность, по которой ученые и рациональные люди сравнивают и оценивают конкурирующие гипотезы) служат признаками хорошей теории и рациональным основанием для ее принятия, по крайней мере, «в научных целях». Но тогда (и это положительный момент в стратегии инструментализма — «дожать» этот вопрос до конца, обратив его против фикционализма) какие еще нужны основания для того, чтобы считать рациональным *принятие* некоторой теории? Если те основания, которые заставляют представителя фикционализма считать материальные объекты и т. п. «полезными фикциями», недостаточны, чтобы признать рациональной веру в «концептуальную систему», предполагающую материальные объекты, то что еще могло бы сделать эту веру рациональной?

Исторически именно этот вопрос разделил представителей фикционализма на два лагеря. Склонный к теологии Дюгем считал, что только томистская метафизика может установить истинность высказываний о реальности, а наука лишь показывает, что некоторые высказывания полезны для предсказания и систематизации данных. С другой стороны, по мнению такого скептика, как Ханс Файхингер, ничем нельзя обосновать реальное существование материальных объектов; мы можем знать только то, что они полезные фикции. Но ни одна из этих позиций не является удовлетворительной. Исследователи, не склонные к католицизму, отказывались видеть в томистской метафизике высший путь к истине по сравнению с современной наукой. Скептицизм же свелся к бесполезному и нелепому утверждению, что раз нельзя дать дедуктивное (или *априорное*) обоснование основным нормам индуктивного исследования, то их следует отбросить. Более того, в скептическом варианте фикционализма есть что-то очень «трогательное», поскольку Файхингер и его последо-

ватели не сомневались в том, что наука дает (приблизительно) правильные предсказания, а стало быть, они принимали индукцию (несмотря на отсутствие ее дедуктивного обоснования), но, с другой стороны, они отказывались верить в то, что в науке возможны *истинные* теории, а значит они отвергали индукцию (или гипотетико-дедуктивный метод, который, как правильно считал Милль, есть наиболее мощный метод индуктивных наук). Почему же мы никогда не можем знать, что научные теории истинны? Потому что, отвечает представитель фикционализма, мы не можем дать никакого дедуктивного доказательства их истинности, даже располагая всеми возможными данными наблюдения. Но мы не можем дать дедуктивного доказательства и тому, что завтра взойдет солнце! Таким образом, представитель фикционализма — непоследовательный скептик; отчасти он принимает индукцию (в том ее качестве, что она дает успешные предсказания опыта), но не полностью (ибо отказывает ей в том, что она дает истинное представление о вещах).

В целом я согласен с инструменталистской стратегией доказательства, но один момент вызывает у меня глубокое беспокойство, а именно та посылка, что целью науки является предсказание опыта (или предсказание плюс «простота», где простота — это некоторая странная цель-в-себе, а не название для большого числа факторов, влияющих на наше суждение о возможной истине). С помощью этой посылки можно легко опровергнуть фикционализм: ведь если не существует различия между *считать, что p и считать, что p дает успешные предсказания* (по крайней мере, когда p — целая концептуальная система), то фикционализм сразу же рушится. Но это все тот же верификационизм, с той лишь разницей, что здесь «единицей значения выступает целостная концептуальная система». Едва ли можно считать, что существует такая вещь, как «цель науки», ибо есть много ученых и много разных целей, и вовсе неверно, что все ученые заинтересованы главным образом в предсказании. Некоторые ученые, например, заинтересованы в открытии определенных фактов относительно радиозвезд, генов, мезонов и т. п. Успешные предсказания им нужны только для подтверждения теорий, а теории нужны отнюдь не для получения предсказаний. Иногда сами по себе предсказания не представляют для них ни малейшего интереса, а если и представляют интерес, то только потому, что с их помощью можно установить истинность или ложность некоторой теории. Сходным образом, и простота не всегда ценится учеными как цель-в-себе; многие ученые заботятся о простоте лишь по той причине и в тех случаях, когда она выступает свидетельством истины. По сути,

единственным значимым различием между следующими утверждениями:

(3) Цель науки — это успешное предсказание

и

(4) Цель некоторых ученых — это знание о том, истинно или нет то, что мезоны ведут себя таким-то образом,

помимо невероятной напыщенности утверждения (3) («цель науки!»), является то, что утверждение (3) выражено в языке наблюдения. Но почему цель науки, если такая вещь вообще существует, или даже цели всех ученых должны формулироваться на языке наблюдения, как и содержание самой науки? Конечно, это требование — просто наследие редукционизма!

Таким образом, фикционализм в целом была отброшен на ошибочных основаниях по той причине, что благодаря верификационизму стало непопулярным разумное и простое различие между истинностью научной теории и истинностью ее наблюдаемых следствий, и тем самым была устранена проблема очевидного разрыва между этими двумя истинами, которая так беспокоила представителей фикционализма. Однако, как мы уже видели, имеется возражение против фикционализма, не затрагивающее редукционистских взглядов на содержание или «цель» науки. Это возражение состоит в том, что если есть рациональные основания принять теорию «в научных целях», то благодаря этим основаниям рациональна и вера в эту теорию, по крайней мере в том смысле, в каком можно «верить» в научную теорию (как в постоянное приближение к истине, но не как в окончательную истину). Фикционализм потерпел крах потому, что он (в противовес утверждаемому Дюгемом) не смог предложить лучшего метода закрепления нашей веры, чем научный метод, и потому что (в противовес утверждаемому Файхингером) отсутствие дедуктивного обоснования научного метода ни в коей мере не доказывает нерациональности его принятия.

Итак, мы рассмотрели возражение против аргумента, ссылающегося на необходимость абстрактных сущностей для науки, — возражение, согласно которому можно с необходимостью считать, что p , но p тем не менее может оказаться ложным. Мы отвергли это возражение не на обычных верификационистских или инструменталистских основаниях, которые, как представляется, опираются на ложную доктрину. Мы отвергли его на другом основании: если у нас

есть причина считать, что p , и мы согласны, что это дает нам основание рационально принять p при любых обстоятельствах, то было бы нелепо затем добавлять: «Но это *недостаточно хорошее* оправдание». Так мог бы добавлять только тот, для кого возможен помимо научного метода высший вненаучный метод; но этот философ, по крайней мере, не заинтересован в *этом*.

IX. НЕУЧТЕННЫЕ СЛОЖНОСТИ

В настоящей статье я углубленно рассмотрел только одну группу вопросов — имеющих отношение к необходимости квантификации по таким абстрактным сущностям, как множества, проигнорировав множество других вопросов. Я мог бы проанализировать еще одну группу вопросов, связанную с существованием так называемых «эквивалентных построений» в математике. Например, числа можно построить на основе *множеств* различными способами. Более того, *множество* не единственное понятие, которое можно выбрать в качестве исходного; мы уже отмечали, что, по крайней мере, предикативная теория множеств взаимопереводима в некотором смысле в утверждения о формулах и истине; и даже непредикативное понятие множества допускает разнообразные эквиваленты: например, вместо определения функций как *множеств* я мог бы определить *множества* как функции. Моя собственная точка зрения состоит в том, что ни один из этих подходов не может считаться «более истинным», чем любой другой; сфера математического факта допускает множество «эквивалентных описаний», но *этому вопросу*, конечно же, можно посвятить отдельную статью.

Кроме того, мы кратко затронули интересный вопрос о конвенционализме. Хотя конвенционалистская позиция никогда не была сформулирована правдоподобным (или даже ясным) образом, она поставила много интересных проблем. Особенно интересен вопрос о том, в какой мере мы можем пересматривать наши основные логические принципы, как мы пересмотрели некоторые из наших основных геометрических принципов в математической физике. Сегодня философы склонны полагать, что сама логика не имеет никакого эмпирического фундамента. Я думаю, это неверно, однако, этот вопрос пришлось оставить в стороне в настоящей статье. Моя цель состояла в том, чтобы дать некоторое представление о многоуровневой сложности проблем, с которыми сталкиваешься при попытке проанализировать хотя бы одну часть философии логики; но, надеюсь, у читателя не сложилось впечатления, что проанализированной мною частью логики все и исчерпывается.

КАК НЕЛЬЗЯ ГОВОРИТЬ О ЗНАЧЕНИИ¹ (Комментарии к статье Дж. Дж. Смарт)

Профессор Дж. Дж. Смарт не только философ, работа которого наводит нас на размышления и побуждает к дискуссии. Он выступает так же в качестве представителя австралийской философии — философии, которая связана с британской и американской философией, но в то же время обладает своими характерными особенностями. Он является в некотором роде философским посланцем, и хорошо, если бы все послы были похожи на него! В своей статье он говорит нам не о своей собственной философской работе, а скорее выражает свои впечатления о положении дел у нас. Как у хорошего философского дипломата, его впечатления полны доброжелательности. Однако, вызывая у нашего гостя дружелюбное отношение, мы можем поставить его в затруднительное положение: достойно сожаления, что я — уроженец этих специфических философских джунглей — чувствую неодолимое стремление отвести его тонкие комплименты и с сожалением возразить, что то, что он превозносит в качестве украшений нашего философского ландшафта, мне представляется лишь сорной травой.

Воззрения, которые излагает Смарт, в основном принадлежат Фейерабенду. Некоторое внимание он уделил также родственным взглядам У. Селларса. В своем комментарии я сосредоточу внимание на позиции Фейерабенда, с которой я знаком лучше, чем с концепцией Селларса. При этом я буду опираться не только на трактовку взглядов Фейерабенда Смарт, но и на ту статью самого Фейерабенда², о которой говорит Смарт. Все ссылки на Фейерабенда будут относиться к этой статье, если не будет указан другой источник.

Редукция. Первая часть статьи Смарт (приблизительно первые семь абзацев) относится к проблеме, являющейся дополнительной по отношению к главной проблеме, и я хочу начать свой анализ именно с нее. Проблема состоит в следующем: как правильное понимать *редукцию* (сведение одной научной теории или дисциплины к другой, например геометрической оптики к электромагнитной теории поля) — как *дедукцию* редуцированной теории из другой теории, к которой она была сведена с помощью редукционных пар,

¹ Перевод выполнен д. ф. н. А. Л. Никифоровым.

² Feysrabend P. K. Explanation, Reduction and Empiricism // Feigl H., Maxwell G. (eds.) Minnesota Studies in the Philosophy of Science, Vol. III. Minneapolis, 1962, p. 28—97.

выражающих «координирующие определения» (то есть с помощью правил соответствия), или же ее следует понимать только как *замещение* одной теории другой? Будет ли старая теория *выводима* из новой теории или она просто замещается последней? Мои комментарии носят двоякий характер: исторический и философский.

Историческое замечание состоит в том, что Фейерабенд и Смарт не знакомы с литературой. Точка зрения Нагеля (редукция посредством редукционных пар) была подвергнута в 1956 году обстоятельной критике Кемени и Оппенгеймом³. Кроме того, в том же серийном издании, в котором опубликована статья Фейерабенда, но томом раньше (том II, 1958) Оппенгейм и я поместили статью, в которой суммировали взгляды Кемени и Оппенгейма. Конечно, вопросы приоритета не имеют большого значения, однако нельзя забывать о литературе по рассматриваемой проблеме.

Мое философское замечание состоит в том, что точка зрения Нагеля не представляется мне столь серьезно ложной. Впрочем, и сам Смарт выражает некоторые сомнения в том, что полемика по этой проблеме является существенной.

Вопрос сводится к следующему: Согласно Смарту, Нагель мог бы «спасти» свою позицию, сказав, что речь идет о *приблизительной* истинности редуцируемой теории (скажем, опять оптики), выводимой из теории, к которой она сводится (скажем, из электромагнитной теории поля). Но Смарт, кажется, считает, что для Нагеля это не было бы *действительным* спасением, так как в защиту Фейерабенда можно было бы (по-видимому, эффективно) возразить: «Фейерабенд, по-видимому, мог бы сказать, что истина не может быть *приблизительной*, *приблизительным* может быть только предсказание. Он, конечно, не будет отрицать полезности старой теории, но при этом, скорее всего, укажет на то, что это больше касается инженеров, чем естествоиспытателей».

Я хочу сказать, что это возражение не является эффективным. Совершенно ясно, что означают слова о том, что теория является *приблизительной истиной*, точно так же как каждый понимает, в каком смысле некоторое уравнение является *приблизительно* правильным: Это означает, что отношения, постулируемые теорией, справедливы не абсолютно точно, а с определенной степенью погрешности. Короче говоря, это означает, что теория не является истинной, однако определенное *логическое следствие* теории, полученное, на-

³ *Kemeny J., Oppenheim P. On Reduction // Philosophical Studies, Vol. 7, 1956, p. 6—19.*

пример, посредством замены выражения «равно» выражением «равно плюс-минус дельта», является истинным⁴. Такого рода логические следствия можно назвать аппроксимированной теорией. Теперь сформулируем следующие вопросы:

(1) Предположим, что O (теория геометрической оптики) не выводима из E (электромагнитной теории поля) с помощью «переходных законов» (bridge laws), но соответствующая аппроксимированная теория O' , таким образом, выводима. В какой степени это изменяет точку зрения Нагеля? *Ответ:* Ни в какой. Сам Нагель вполне осознает, что не подвергаясь модификации редуцируемая теория O несовместима с теорией E , к которой она редуцируется.

(2) Может ли понятие «луч света» быть введено в теорию E таким способом, чтобы соответствующая теория O' оказалась корректной? *Ответ:* Да. (Определим «луч света» как нормаль к фронту волны.)

Если учесть (1) и (2), то я не вижу, из-за чего, собственно, поднимать шум. Неспособность Фейерабенда увидеть эти очевидные вещи на первый взгляд объясняется просто его упрямством. Однако позднее мы покажем, что это не случайный факт для позиции Фейерабенда. В этом выражается его стремление свести к минимуму и даже вовсе отвергнуть представление о науке как кумулятивном процессе накопления знания о природе, для того чтобы увеличить правдоподобность его собственной странной точки зрения, согласно которой самое большее, на что мы можем надеяться, — это полу-

⁴ Замена «равно» на «равно плюс-минус дельта» приведена только в качестве примера. Другими возможными заменами могли бы быть следующие: (а) замена «равно» на «равно с вероятностью, большей чем 90%», и (б) замена «равно» на «равно при условии S ». Конечно, все эти (а также другие) виды замены могут осуществляться одновременно. В результате этого в каждом случае происходит *ослабление* первоначальной теории, понимаемое в логическом смысле слова «ослабление», то есть получающаяся в результате замены теория всегда является *логическим следствием* первоначальной теории. Используемый в тексте термин «аппроксимированная теория» может быть естественно расширен на все эти виды замен. Я утверждаю: Несмотря на тот факт, что научные постулаты обычно являются ложными, научное познание развивается кумулятивно в том смысле, что, даже если хорошая научная теория не является в *точности* правильной, разумно аппроксимированная теория обычно является правильной и в конечном итоге оказывается устойчивой частью наших «основ познания». Действительно, если бы научное познание не было кумулятивным в этом смысле, то едва ли можно было бы понять, почему оно должно иметь какое-либо теоретическое (а не только инженерное) значение.

чить не корректное объяснение явлений, а различные их альтернативные объяснения посредством ложных теорий. Основное содержание статей Фейерабенда сводится к тому, что мы должны стремиться к плюрализму теорий, а не к единственному правильному объяснению того или иного явления, и мы должны ожидать, что все эти теории впоследствии окажутся ложными и будут вытеснены новым рядом ложных теорий, не «соединяясь» в то, что Фейерабэнд насмешливо называет «единственной истинной теорией». (Это может удивить того читателя, который со взглядами Фейерабенда познакомился лишь по изложению их Смартom; однако Сمارт говорит только о тех воззрениях Фейерабенда, которые ему представляются особенно привлекательными и интересными.)

Отметим, между прочим, что в отличие от Фейерабенда Кемени и Оппенгейм признавали «редукцию посредством редукционных пар» (редукцию Нагеля) важным специальным типом редукции.

Значение и здравый смысл. В изложении Смартa наиболее интересная и основная часть концепции Фейерабенда основывается на отождествлении значения термина с определенной принятой теорией, содержащей этот термин. Применяя эту точку зрения к терминам обычного языка, Фейерабэнд приходит к выводу, что концептуальная схема нашего здравого смысла является ложной теорией. Следует подчеркнуть, что отождествление значения с теорией в статье Фейерабенда проводится не открыто, а завуалированно. Фейерабэнд не *говорит*, что теория является значением термина. Он постепенно переходит от термина «значение» к такому выражению, как «принятое употребление» (которое является в некотором смысле достаточно невинным), а затем ссылается на эмпирические убеждения, содержащие термин «значение», как на примеры «принятого употребления». Это повторяется в статье неоднократно. Например, предполагается, что «способ», которым термины «вверх» — «вниз» «употреблялись» Фалесом, показывает, что «космологические предположения были «скрыто включены в общие для того времени средства выражения» (Фейерабэнд П., указ. соч., с. 85). (Эти «космологические предположения» в том виде, как они сформулированы Фейерабэндом, включают в анахроническом выражении ньютоновское понятие «сила».) Рассматривая этот пример, Фейерабэнд просто *спутал* «употребление слов» (в лингвистическом смысле) с *вхождением* их в некоторый контекст, и вся его аргументация основывается на этой грубой ошибке.

Как только такое отождествление принято, связь различных частей доктрины Фейерабенда становится ясной. Основное ее содер-

жание — в том виде, как она представлена и интерпретирована Смарттом, — состоит в том, что если один и тот же термин встречается в двух различных теориях T_1 и T_2 , то нельзя предполагать, что он имеет одно и то же значение. Для Фейерабенда это, в сущности, является тавтологией: фраза «Различие теорий влечет различие в значении» выражает у него только то, что «Различие теорий влечет различие теорий». Фейерабенд сам утверждает, что «Значения не являются инвариантными относительно процесса объяснения», то есть объясняемый факт содержит термины, которые изменяют свое значение, когда рассматриваемое утверждение логически выводится из данной теории.

Самая радикальная мысль Фейерабенда (Указ. соч., с. 90—91) состоит в том, что поскольку значение (в фейерабендовском понимании термина «значение») психологических терминов обычного языка (например, «ощущение», «боль») является некоторой теорией, а эта теория может быть ложной, то мы в действительности не должны беспокоиться по поводу «лингвистических» возражений материализму⁶, «относительно которых можно показать, что они совершенно иррелевантны». «Ментальные коннотации теоретических терминов могут быть ложными».

Так как Фейерабенд не проводит различия между концептуально необходимыми суждениями (которые, согласно его взгляду, основываются на эмпирических теориях) и эмпирическими высказываниями, то для него, например, фраза: «Если я утверждаю, что я испытываю боль, и мне ясно, что я утверждаю, и если я говорю это искренне, то я должен испытывать боль» была бы высказыванием, которое будущая эмпирическая теория может так же опровергнуть, как и любое другое утверждение. Я отдаю себе отчет в том, что су-

⁶ Под «материализмом» я здесь понимаю доктрину, согласно которой боль, гнев и т. п. являются нейрофизиологическими состояниями (или событиями). В одной своей работе я утверждал (Putnam H. The mental life of some machines), что эта доктрина некорректна (по «лингвистическим» основаниям, что Фейерабенд, несомненно, оценил бы как иррелевантный аргумент). Я не отрицаю «материализма» в более широком смысле — как воззрение, рассматривающее человеческое существо в целом как физико-химическую систему, имеющую некоторую кибернетическую организацию. Действительно, я думаю, что в этом более широком смысле материализм является правильным. И я согласился бы с Фейерабендом в отрицании «лингвистической» аргументации против материализма в этом более широком смысле. Однако я не знаю никакой аргументации такого рода, которая заслуживала бы серьезного рассмотрения.

ществуют трудности и в обосновании дихотомии аналитического и синтетического, и в обосновании того, что приведенное утверждение является аналитическим. Однако несомненно, что легкий путь, избранный Фейерабендом в отношении «лингвистической» аргументации, является *слишком* легким.

Если мы отвергаем истолкование Фейерабендом «значения» (что мы должны сделать, если вообще хотим говорить о значении в обычном смысле), то мы вполне можем сказать, что, даже если *некоторые* убеждения здравого смысла относительно боли и являются *ложными*, все-таки, несомненно, не ложно то, что *существует* боль — не в «пиквикском» смысле слова «боль», а в обычном смысле. И если некоторые утверждения о боли концептуально необходимы (то есть необходимы благодаря правилам языка, считая, что эти правила *не* предполагают в сущности некоторой «ложной теории»), то любое определение боли — материалистическое или какое-либо другое — должно быть совместимо с этими утверждениями.

В свою очередь Фейерабенд должен или (1) совершенно отказаться от различения аналитического и синтетического, но тогда он вообще не имеет права использовать термин «значение», или (2) утверждать, что *правила обыденного языка* (а не просто некоторые эмпирические убеждения здравого смысла) в отношении слова «боль», в сущности, предполагают некоторую ложную теорию. В этом последнем случае он должен сказать, чем является эта теория, и показать как то, что она *является* эмпирической⁶, так и то, что она действительно предполагается правилами нашего обычного языка. Однако это привело бы к обсуждению именно лингвистических вопросов, которые он склонен объявлять иррелевантными.

Примеры Фейерабенда и Смарта. Следует отметить, что примеры, приводимые Фейерабендом и Смартom, с лингвистической точки зрения неубедительны. Согласно Фейерабенду, под словом «температура» мы понимаем не то, что понимал Галилей (что он подразумевал под соответствующим итальянским словом): Фейерабенд объясняет это тем, что мы отвергли высказывание «Температура, фиксируемая термометром, не зависит от химического состава исполь-

⁶ По-видимому, Фейерабенд считает, что ложность рассматриваемой теории состоит в том, что боль и т. д. считаются *мыслимыми*. Тогда возможно, что, скажем, «Гнев есть состояние ума» является семантическим следствием правил обыденного языка (или, проще говоря, что оно является аналитическим). Но почему это является (ложной) *эмпирической теорией*? Почему это нельзя рассматривать как некоторую тавтологию, которая просто иллюстрирует наше использование словесного оборота «состояние ума»?

зуюмой жидкости», которое Фейерабенд считает существенным для галилеевского понимания. Если бы это высказывание действительно было «включено» в галилеевское понятие температуры, то Галилей не в состоянии был бы понять отрицание данного утверждения. Это, конечно, абсурдно. То, что действительно понимал под температурой Галилей, было внутренним свойством тела, которое измеряется термометром, а не результат измерения. И Галилей мог бы *понять* утверждение о том, что измеренная температура не вполне точно соответствует подлинной температуре и что измеренная температура зависит в некоторой степени от используемой в термометре жидкости, точно так же, как поняли бы его вы или я независимо от уровня нашей компетенции в физике.

Все остальные примеры Фейерабенда не более убедительны. Если бы существенной составной частью значения слова «вниз» было то, что ничем не поддерживаемая Земля упала бы вниз (или что люди провалились бы к антиподам) и что «вниз» всегда означает одно и то же направление, то древние не могли бы делать умозрительных утверждений, что Земля является ничем не поддерживаемым шаром (и о том, что антиподы с нее не падают). Мы не можем говорить, что «Некоторые холостяки в действительности могут быть женаты», так как «быть неженатым» мы включаем в понятие «холостяк»; точно так же мы не можем говорить, что неподдерживаемый объект может не падать вниз или что он может не всегда падать в одном и том же направлении, если противоположное утверждение является лингвистическим требованием. Однако не существует ни малейших доказательств того, что утверждение «Неподдерживаемый объект упадет вниз» когда-нибудь было лингвистическим требованием в каком-либо смысле. И не существует доказательств того, что высказывание «Вниз — всегда является одним и тем же направлением» было когда-либо лингвистическим требованием. Всегда существовали волшебные рассказы, в которых неподдерживаемые объекты не падали вниз, и рассказчики — даже маленькие дети — никогда не затрудняются в понимании этих сказок. Теории шарообразности Земли существовали в течение тысячелетий. Фейерабенд ошибается именно потому, что такие теории можно объяснить кому-угодно, даже ребенку, на совершенно обычном языке.

Некоторые примеры Смарта даже еще менее убедительны. То, что выражение «дымовая труба» изменило свое значение в результате развития современной физики, явно ложно.

Злоупотребление философами термином «значение». На мой взгляд, очевидно, что Фейерабенд неправильно употребляет

термин «значение». К тому же он не одинок в этом: В последние тридцать лет неправильное употребление термина «значение» было одним из наиболее распространенных, хотя и наименее успешных способов «обоснования» философских высказываний. Как же возникло такое печальное положение дел?

Вина за это должна быть возложена на логический позитивизм. «Верификационная теория значения» («значением предложения является метод его верификации») сначала была не чем иным, как убедительным переопределением смысла термина «значение». Если бы характеристика метафизических высказываний как «бессмысленных» означала бы только то, что эти высказывания эмпирически непроверяемы, то это было бы вполне безвредно (метафизики всегда *говорили*, что их утверждения не являются ни тавтологиями, ни эмпирически проверяемыми утверждениями). Однако эта характеристика не столь безобидна, как может показаться, поскольку позитивист надеется на то, что мы признаем его переопределение термина «значение», *сохраняя в то же время уничтожительный смысл выражения «бессмысленный» в обычном (лингвистическом) понимании, то есть в буквальном значении — не иметь смысла.* Поскольку же тезис, не имеющий смысла, нельзя ни обсуждать, ни оспаривать, постольку человек, утверждающий, что метафизическая философия — «бессмысленна», чувствует, что он вправе отрицать ее в целом, не приводя никаких аргументов против нее. Если утверждение «бог существует» бессмысленно, то тщательное рассмотрение аргументов «за» и «против» этого утверждения, предпринятое Юмом в «Диалогах о естественной религии», необязательно; действительно, как могут существовать аргументы «за» или «против» *бессмысленного предложения?*

Позитивистский ответ на этот вопрос состоит в том, что верификационная теория значения является «экспликацией», а экспликация не обязательно должна точно соответствовать «доаналитическому использованию» экспликандума. Однако это нечестный ответ. «Метод верификации» не является экспликацией понятия значения, которое используется лингвистической теорией и в повседневной жизни, и в действительности (как мы видели) не предназначался служить в качестве таковой. Для достижения своей цели (устранение метафизики, нормативной этики, теологии и т. д.) позитивизму было необходимо, чтобы верификационная теория значения как раз и не была экспликатом обычного понятия «значения».

Одно из наиболее широко обсуждаемых возражений против верификационной теории значения в ее первоначальной форме («зна-

чение предложения есть метод его верификации») принадлежит Куайну. Это возражение состоит в том, что метод верификации служит для проверки не изолированного *предложения*, а *целой теории*. В процесс эмпирической проверки включается большая группа предложений, а не отдельное предложение. В самом деле, мы можем рассматривать проверку как проверку одного-единственного предложения S только в том случае, если *другие* предложения S_1, S_2, \dots, S_n , включенные в контекст данной теории, (практически) не вызывают сомнений. Но в таком случае эта проверка является «проверкой S » благодаря значению не самого S , а благодаря значению S вместе со значениями S_1, S_2, \dots, S_n и тому факту, что S_1, S_2, \dots, S_n рассматриваются как эмпирически истинные.

На эту трудность логические позитивисты реагировали двумя разными способами. Некоторые из них согласились с Куайном и отказались от тождества значения и изменения значения. С их точки зрения, существует такая вещь, как осмысленность (и ее критерием является включенность соответствующего предложения в эмпирическую теорию), и вопросы верификации косвенно входят в установление того, является ли осмысленным данное слово или предложение. Однако, с их точки зрения, нет вообще никакого смысла в вопросе о том, имеют ли два слова одно и то же значение или изменило ли некоторое слово свое значение. Конечно, мы можем спросить, имеют ли определенные два слова один и тот же *экстенционал*, однако этот вопрос имеет фактуальный характер и он должен быть решен с помощью обычных научных процедур. (Для этой точки зрения является сложным вопрос о том, какие обычные научные процедуры должны быть использованы, чтобы решить, являются ли все холостяки неженатыми или нет; отчасти неопределенный ответ заключается в том, что такие утверждения «оправдываются их местом в некоторой системе».)

Другой подход, развиваемый Карнапом, состоит в следующем: Вопросы верификации (то есть существования проверяемых следствий) входят в установление того, является ли некоторое множество предложений эмпирической теорией или не является таковой. Роль терминов в формулировке эмпирических теорий — вот что делает их осмысленными (как и в предыдущем подходе). Таким образом, вопросы верифицируемости снова косвенно входят в проблему осмысленности, но не в вопросы *тождества* или *различия* значений. Последние касаются *классов синонимичности*, к которым принадлежит слово, и решаются с помощью обращения к семантическим правилам языка. Таким образом, фактически Карнап оперирует *двумя* понятиями

ми значения: Он использует обычное (лингвистическое) понятие значения, когда речь идет о *синонимии*; кроме того, он использует верификационную теорию значения в своей современной, значительно модифицированной форме, когда обсуждает вопрос: «Включается ли вообще некоторый термин в язык?».

Фейерабенд в значительной степени не соглашается с логическим позитивизмом, особенно с его современной формой. Однако неправильное употребление им термина «значение», по-видимому, было вызвано верификационной теорией значения. Так, в одном месте он утверждает, что Галилей под температурой имел в виду «то, что измеряет термометр», и в этом утверждении содержится зародыш истины, но термин «измеряет» он понимает в довольно сомнительном смысле, превращая Галилея в инструменталиста ⁷. В другом месте он использует положение Нагеля, что «процедуры» феноменологической термодинамики фиксируют значение термина «температура». Прав я или нет, предполагая, что именно позитивизм побудил Фейерабенда к небрежному обращению с термином «значение», но факт остается фактом: Фейерабенд не может избежать большинства трудностей, с которыми столкнулись позитивисты (или, скорее, он не мог бы избежать их, если бы попытался когда-нибудь подвергнуть свое собственное использование в философии термина «значение» хоть сколько-нибудь внимательному анализу).

Для того чтобы увидеть это, достаточно вспомнить, что для Фейерабенда значение некоторого термина зависит от всей *теории*, содержащей этот термин. Поэтому он не может уйти от следующих вопросов: *Какие* теории порождают значения терминов? Поскольку теория содержит много различных терминов, то как мы можем сказать, значение каких *терминов* она детерминирует? Какие изменения в теории или, более широко, в эмпирических убеждениях *не*

⁷ Фейерабенд утверждает, что термометр «измеряет» не температуру, а сложную функцию многих физических переменных (включая температуру). Он подразумевает при этом, конечно, то, что *точное показание* термометра связано со сложной функцией, состоящей из многих переменных. При подобном использовании слова «измеряет» получается, что каждый измерительный инструмент является *совершенным* и только мы иногда ошибаемся относительно того, что измеряется. Приписать Галилею точку зрения, что «температура есть то, что измеряет термометр», где слово «измеряет» берется именно в этом смысле, — значит превратить Галилея в крайнего операционалиста. Излишне говорить, что это неразумное использование слова «измеряет», и неразумно говорить, как это делает Фейерабенд, что сегодня мы не верим в то, что термометры измеряют температуру.

влиают на значение терминов? В статье Фейерабенда не предложено никаких ответов на эти вопросы, а ведь эти вопросы выражают как раз те трудности, которые заставили защитников верификационной теории значения ограничить свою теорию вопросами *смысла* и искать понимание *синонимии* где-то в другом месте, если оно вообще существует.

Можно, конечно, избрать крайнее направление и утверждать, что *любое* изменение теории является изменением значения терминов. (Каких терминов? Всех? Даже логических связей? А почему бы и нет?) Однако я надеюсь, что Фейерабэнд не хотел бы принимать такого истолкования этой проблемы. Так как сказать, что *любое* изменение в наших эмпирических убеждениях относительно множества X -ов приводит к изменению значения термина X ⁸, значило бы отрицать различие между вопросами о значении и вопросами о факте. Сказать, что семантические правила английского языка вообще нельзя отличить от эмпирических убеждений людей, говорящих по-английски, значило бы выбросить за борт понятие семантического правила английского языка.

Любопытно, что Фейерабэнд *не* следует этим путем. Действительно, многие его цели были бы достигнуты быстрее, если бы он предпочел следовать Куайну и совершенно отбросил бы теорию значения. Он хочет показать, что лингвистическая философия irrelevantна, и вводит в заблуждение: ясно, что если она опирает на ошибочное понятие о том, что существуют такие вещи, как *правила языка*, то она глубоко заблуждается. (Такова вся наука лингвистика; это возражение Куайн пытался обойти в работе «Слово и объект»⁹ посредством допущения некоторых построений лингвистических описаний естественного языка, причем эти описания должны были удовлетворять некоторым операциональным и теоретическим условиям.) Он стремится показать, что если в ответ на фразу «Боль есть некоторый процесс в коре головного мозга» говорят, что это только изменение значения слова «боль», то это возражение неубедительно. Если «изменение значения» само является *бесмысленным* понятием, то его цель достигнута. Он хочет показать, что в основе ложных теорий лежит обычный язык; если нельзя провести различия между

⁸ Второе вхождение « X » в предложение текста следовало бы поместить в квазикавычки («уголки» Куайна) для того, чтобы избежать ошибки смешения упоминания и использования. В настоящей статье я игнорирую эти логические тонкости.

⁹ Quine W. V. O. Word and Object. Cambridge, Mass., 1960.

«обыденным языком» и «здравым смыслом» (то есть повседневными убеждениями людей, говорящих на обыденном языке), то это означает, что большинство людей верят во многие ложные вещи, но кто когда-либо сомневался в *этом*?

Однако вся сенсационность исчезла бы, если бы Фейерабенд следовал этим путем, так как «сенсационность» здесь обусловлена скрытым переходом от необычного понимания значения к его обычному пониманию и обратно. Рассмотрим, например, утверждение, вызвавшее интерес Смарта, что даже термин «дымовая труба» изменил свое значение, так как мы отказались от точки зрения здравого смысла при понимании внутреннего строения труб. Если «значение» есть точка зрения здравого смысла, то это тавтология: «Мы изменили нашу точку зрения, так как мы изменили нашу точку зрения». Только благодаря незаметному переходу от того факта, что понятие «значение» в *фейерабендовском* смысле (если таковое действительно существует) изменилось, к утверждению о том, что изменилось значение в *обычном смысле*, получается «сенсационный» результат.

И наконец, я затрону одно из главных утверждений Фейерабенда, которого в некотором смысле касается вся статья Смарта и которое просто не могло бы быть выведено без этого «перехода». Это утверждение состоит в том, что, поскольку повседневный язык предполагает ложные теории, мы можем *отбросить повседневный язык* (с. 76). Нет ничего удивительного в том, что мы можем изменить многие убеждения нашего здравого смысла; мы можем высказать различные мысли на обычном английском, французском, немецком и т. д. языках; эти мнения являются «значениями» наших терминов в *фейерабендовском* смысле слова «значение»; но, только обосновав то, что они являются также «значениями» в обычном смысле, мы приходим к заключению, что можно всецело отбросить повседневный английский, французский, немецкий и т. д. языки. Конечно, здесь Куайн и Фейерабенд разошлись бы друг с другом. Если «обычное использование» термина X зависит от ложного суждения, что все X являются P , то Куайн рекомендовал бы заменить его истинным суждением, а именно: «Некоторые X не есть P », а не отбрасывать обязательно термин X . Если «здравый смысл» ложен, то следует выяснить, что же истинно. Можно ввести много специальных терминов, но что обычный, не специальный язык нельзя использовать для того, чтобы высказать истину, что он *существенно поражен ложью*, — это заключение, которое Куайн отверг бы, как он отвергает «эссенциализм» вообще. Действительно, взгляды Фейерабенда представляют собой беспорядочную смесь куайновского неже-

лания отделять значение от эмпирической теории и того эссенциализма, с которым борется Куайн.

Что есть значение? Я неоднократно говорил об «обычном смысле» слова «значение». В чем он заключается?

В некотором отношении мы все достаточно хорошо знаем, что означает слово «значение». Поэтому нам не хватает *не синонима* для слова «значение» (например, слова «смысл»), а концептуального анализа. Попытки концептуального анализа понятия значения иногда принимают форму исследований оснований лингвистической теории (ср. серьезную книгу Пауля Зиффа¹⁰), а иногда форму формальных моделей. Я не собираюсь здесь претендовать на что-либо подобное. Надеюсь, что мои утверждения будут приемлемы для тех, кто не отрицает, подобно Куайну, теорию значения целиком.

Теория значения опирается на ту идею, что естественный язык имеет правила. (Зифф возражает против термина «правило», но его «структурные регулярности» и «структурные планы» являются «правилами» в моем использовании этого термина.) Эти правила являются иногда синтаксическими, а иногда несинтаксическими. Синтаксические правила детерминируют, какие звуки суть звуки фонем языка, какие последовательности фонем — морфемы (более строго, алломорфы морфем) и какие последовательности морфем — правильно построенные предложения. Несинтаксические правила отличают лингвистически *правильное* использование предложений от лингвистически *неправильного* использования. Назвать женатого человека «холостяком» или сказать, что геометрическая точка «грациозна», — значит нарушить семантические правила английского языка.

Основная задача семантической теории состоит в том, чтобы исследовать различные виды несинтаксических правил, существующих в языке, и попытаться ясно их сформулировать. Я ограничусь приведением немногих примеров, выражающих форму и содержание семантических правил естественного языка. Однако даже на этом первом этапе рассмотрения некоторые вещи кажутся ясными. Все согласится с тем, что знание правил родного языка у людей, пользующихся этим языком, является неявным; только весьма искусственный человек может выразить (некоторые) *синтаксические* правила своего родного языка. Бессознательный и обязательный характер лингвистических правил был замечен давно. «Значение» некоторого слова является функцией правил, управляющих его употреблением. Эти правила, помимо всего прочего, детерминируют, какие выражения

¹⁰ Ziff P. *Semantical Analysis*. Ithaca, 1960

синонимичны (то есть совпадают, согласно лингвистическому соглашению), какие выражения имеют более одного значения (то есть управляются альтернативными наборами правил), какие предложения и при каком их истолковании являются аналитическими и т. д.

Бессознательный характер лингвистических правил важен для понимания того, что происходит, когда спрашивают о «значении» некоторого слова. Спрашивающий хочет получить знание правил, управляющих употреблением данного слова, для того, чтобы он сам также смог его употреблять. Однако он хочет получить не явную формулировку этих правил, а некоторый вид имплицитного знания, о котором я говорил выше. Таким образом, в данной ситуации задача отвечающего состоит в том, чтобы сказать нечто такое, из чего спрашивающий, опираясь на свое (неявное) лингвистическое знание, может «выбрать» ту информацию, которую он желал получить, — «информацию», которую ни спрашивающий, ни отвечающий не могут выразить словесно, и которая «добывается» посредством процесса, которого сегодня *никто* не понимает.

Изложенное выше может быть принято в качестве ответа на вопрос: «Что является значением такого-то и такого слова?» Традиционные определения в словарях обычно содержат фантастическую смесь эмпирической и лингвистической информации. Тем не менее они часто позволяют говорящим вполне овладеть использованием слов, которых они раньше не понимали.

Например, предположим, что человека, с рождения говорящего на английском языке, спрашивают о значении слова «gold» («золото»). Вероятно, что он даст спрашивающему массу эмпирической информации о золоте (что оно драгоценно, имеет обычно желтый цвет, не подвержено коррозии и т. д.) в дополнение к существенной лингвистической информации о том, что «золото» есть наименование некоторого металла. И даже если золото станет «дешевым, как грязь», начнет ржаветь или окажется зеленым, значение слова «золото» не изменится. Только в том случае, если бы мы прекратили использовать «золото» как наименование некоторого металла или употребили бы это слово для названия другого металла, основное значение этого слова изменилось бы. (Я говорю «основное значение», поскольку *сопутствующие значения* слова «золото» действительно *зависят* от того, что фактически золото обычно желтого цвета, драгоценно и т. д.)

Следует заметить также, что можно знать значение слова «золото», не зная, как *объяснить*, что данная вещь является или не является золотой. (Конечно, в этом случае человек должен быть спосо-

бен *установить*, что он имеет дело с золотом, иначе это слово не могло бы использоваться как наименование реального металла. Однако неверно, что только те люди, которые могут отличить золото от других металлов, знают значение этого слова или знают это значение *лучше*. Просто они знают *больше о золоте*.)

То же самое относится к примеру Фейерабенда со словом «температура». Поскольку мы продолжаем использовать слово «температура» для того, чтобы относить его к одной и той же физической величине, постольку мы не скажем, что «значение» этого слова изменилось, даже если мы неоднократно пересматриваем наши представления относительно точных законов для этой величины и независимо от того, насколько изощренными могут стать наши инструменты для измерения температуры. Однако некоторая «теория» относительно значения слова «температура» все же необходима. Суть этой теории заключается в том, что величина, которую мы определяем как «температура» и которая количественно измеряется посредством термометром или еще каким-либо образом, является величиной, большая или меньшая степень интенсивности которой измеряется человеческой сенсорной способностью как *более теплое* или *более холодное*. Это не означает, что чувства человека никогда его не обманывают, но когда они не обманывают, когда различия в ощущениях тепла обусловлены различиями в некотором свойстве объекта, а не субъекта, то именно различия в «температуре» являются причиной этого различия в ощущениях. Использование слова «температура» основывается на том эмпирическом факте, что существует отдельная физическая величина (фактически молекулярная энергия), которая, как правило, лежит в основе различий в «ощущении тепла», поэтому утверждение, что «если X имеет более высокую температуру, чем Y , то X теплее, чем Y », является аналитическим, то есть слова «температура» и «теплее» связаны семантически, и, следовательно, человеческая чувственность лежит в основе «стимулирующего значения»¹¹ слова «теплее». Слово «температура» является «теоретически нагруженным», и, к счастью, эта теория правильна.

В предшествующем разделе я утверждал, что Галилей использовал определенное слово (или, скорее, соответствующее слово итальянского языка) для обозначения физической величины, которую мы сегодня называем «температура». Но на каком основании можно так говорить? Одна причина, конечно, состоит в том, что все ученые

¹¹ Термин «стимулирующее значение» («stimulus meaning») взят из Quine W. V. O. *Word and Object*. Cambridge, Mass., 1960.

признают хорошо известную редукцию феноменологической термодинамики к статистической механике (отождествляющей «температуру» со средней кинетической энергией молекул). Это «редукция» в смысле Нагеля — редукция посредством редукционных пар, — и она приводит к дедуктивному вышеданию из статистической механики очень хорошо аппроксимированной теории феноменологической термодинамики. Признание этой редукции означает признание того, что феноменологическая термодинамика и статистическая механика (с добавленным к ней постулатом «температура есть средняя кинетическая энергия молекул») являются теориями «температуры» в одном и том же смысле, как справедливо отметил это Нагель.

С этим примером связаны два вопроса, носящие больше «лингвистический» характер:

(1) Опираясь на *лингвистическую интуицию*, лингвистически более правильно говорить, что Галилей измерял и рассуждал о величине, которую мы называем «температурой», но наши представления относительно этой величины несколько отличаются от тех, которые были у Галилея. Тот факт, что говорящие на родном языке имеют определенную интуицию правильности, конечно, ничего не *доказывает*, но этот факт *является* частью данных, имеющих в распоряжении лингвиста. Когда кто-нибудь выступает против широко распространенной лингвистической интуиции, он должен иметь для этого веские причины.

Фейерабенд, как можно заключить из его позиции, в целом отрицает этот вид обращения к лингвистической интуиции. Для него «лингвистическая интуиция» говорящих на родном языке просто отражает ложные теории, которые они сделали своими внутренними убеждениями. Таким образом, подчинение лингвистической интуиции свидетельствует о подчинении ложным теориям и, следовательно, неприемлемо¹². Но здесь имеется некоторая путаница. Лингвисты и специалисты в области философии языка опираются на лингвистическую интуицию, а также на методологические соображения (например, простоту полного описания синтаксиса и семантики языка) при исследовании корректной семантической характеристики некоторого слова *X*. Утверждая, что то-то и то-то есть корректная семантическая характеристика слова *X*, они не столько заботятся о существовании *X*-ов, сколько о корректности тех эмпирических мнений, которые могут вытекать из обозначения чего-либо с помощью *X*.

¹² Ср.: *Feyerabend P. K. How to Be a Good Empirist // Baumrin B. (ed.) Philosophy of Science (The Delaware Seminar, Vol. II). N. Y., 1963, p. 36.*

Если использование слова *X* действительно предполагает ту или иную ложную эмпирическую теорию, то мы могли бы надеяться обнаружить ее именно с помощью этой процедуры. Если мы отвергаем лингвистическую интуицию прежде всего как нечто «грязное» (что существенно заражено ложью), то мы никогда не раскроем семантической характеристики ни одного слова и, следовательно, никогда не обнаружим, как это утверждает Фейерабенд, что большинство слов предполагает ложные эмпирические теории, что различные термины изменили свои значения, что значения не инварианты в процессах редукции и объяснения, то есть никогда не установим ложность или истинность тезисов, подобных тезисам Фейерабенда.

(2) Если даны два семантических описания языка — всего или его части (даже если они приблизительно равны по простоте и фактуальной адекватности), всегда существует строгое методологическое основание для того, чтобы предпочесть то из них, которое постулирует меньшее изменение значения. А именно тот факт, что два разных употребления терминов языка подчиняются *одной и той же лексической* характеристике (обладает такими-то и такими общими особенностями), является фактом относительно языка; в то же время то, что они могут иметь два различных вида «записи», является тривиальным фактом, так как значение любого слова может «расщепиться» на сколь угодно много направлений. Таким образом, нельзя обнаружить семантическую структуру языка, не признавая максимы: «Не постулировать различия значений без необходимости». (Зифф называет это «ластиком Оккама».)

Таким образом, мы снова видим, что если имеются операциональные и методологические ограничения, соответствующие лингвистической теории, то неразумно говорить, что такие слова, как «температура», изменяют свое значение всякий раз, когда мы изменяем наши теории относительно соответствующих величин. Мы привели два примера такого рода ограничения, которые признаются семантиками. Куайн скептически относится к тому, что такие ограничения выбирают *единственное* семантическое описание естественного языка. Однако в данный момент это нас не тревожит ¹³.

¹³ Между прочим, ограничения, принимаемые Куайном, являются, очевидно, слишком слабыми. Например, по формальным основаниям Куайн требует только, чтобы схема перевода была в целом рекурсивной. Это приводит к признанию в качестве «корректных» удивительно неправильных переводов. Практически нечто напоминающее замену одного слова другим используется всегда, по крайней мере на первом этапе. По операциональным основаниям Куайн отвергает подчинение лингвистической интуиции говорящих на род-

Наша сегодняшняя трудность состоит не в том, что перед нами находится множество «равноценных и обоснованных описаний» английского языка и мы не знаем, какое из них выбрать (или не знаем, имеет ли смысл вообще выбирать), а в том, что не построено даже *одного* вполне адекватного описания английского языка. И естественно, мы не имеем дело с ситуацией, при которой *каждое* описание является адекватным (что допускает Куайн в работе «Слово и объект»).

Обычный смысл слова «значение» может быть несколько неопределенным, но большой интерес лингвистов, а также представителей философии языка к семантической теории и проявление интеграции семантических и синтаксических исследований в лингвистике позволяют надеяться на то, что он может быть прояснен и дополнен рядом более специальных понятий. Мы находимся на пути познания обоснованного и теоретически плодотворного способа говорить о «тождестве значений» и об «изменении значения» в лингвистическом смысле. В защиту Фейерабенда Смарт выдвигает предположение, что Фейерабэнд, возможно, говорит о «значении» *в некотором ином смысле*. Но каков *точно* этот другой смысл? Какие ограничения накладываются на этот иной способ говорить о «значении»? В другой статье Фейерабэнд говорит, что как раз («эмпирическое») значение» и не остается инвариантным. Единственной существующей теорией «эмпирического значения» является верификационная теория значения, и, как мы видели, в своей современной форме эта теория дает только критерий эмпирической осмысленности, а не критерий эмпирической *синонимии* (тождества «эмпирических значений»). Я не могу присоединиться к Смарту и вместе с ним восхвалять как важное открытие то, что термины постоянно изменяют свое значение в некотором совершенно неопределенном смысле слова «значение»; я не могу также согласиться с утверждением, что поскольку обычный язык включает в себе ложные теории, причем совершенно непонятным образом, постольку обычный язык должен быть отброшен.

ном языке как ограничение по той причине, что наличие этого ограничения заставляет считать философские вопросы решенными. (Откуда возникает эта «лингвистическая интуиция», если не из имплицитной теории языка, спрашивает Куайн. И какие были ограничения на эту имплицитную теорию?) Но это выглядит пуризмом. Куайн прав в том, что *основания* лингвистической теории не могут постоянно опираться на апелляцию к лингвистической интуиции (как это делают, например, Хомский и его последователи), но это не означает, что мы вообще не должны прибегать к таким апелляциям. К счастью, мы вполне можем довольствоваться науками, концептуальные основания которых еще далеки от нормального состояния.

Язык — это первая обширная область человеческих познавательных способностей, для которой мы начинаем получать не слишком упрощенное описание. Благодаря работам современных трансформационных лингвистов¹, по крайней мере некоторым языкам человечества дано весьма совершенное описание. Часть характеристик этих языков представляются *универсальными*. Если эти характеристики являются «специфическими для человеческого рода» — «не объяснимыми на общих основаниях функциональной полезности или простоты, применимых к любым произвольно взятым системам, выполняющим функции языка», то они могут пролить свет на структуру сознания. Хотя невероятно трудно сказать, в какой мере выявленная таким образом структура окажется универсальной структурой языка, а не универсальной структурой врожденных общих стратегий обучения², сам факт такого обсуждения свидетельствует о богатстве и универсальности поставляемого лингвистами материала, равно как и о глубине анализа, поскольку кандидатами на роль «специфических для человеческого рода» выступают не поверхностные или феноменологические характеристики языка, а характеристики, выявляемые на уровне глубинной структуры.

В философском плане наиболее серьезным недостатком этого анализа является то, что он не затрагивает значения слов. Анализ глубинной структуры лингвистических форм дает нам несравнимо более мощное описание *синтаксиса* естественных языков, чем мы имели раньше. Однако аспект языка, связанный со словом «значение», несмотря на обилие героических, но неудачных попыток, по-прежнему остается в тени.

В своей работе я хотел бы выяснить, почему это так и должно быть. С моей точки зрения, причина, по которой так называемая се-

* The Meaning of "Meaning". Впервые опубликована в кн.: *Gunderson K.* (ed.) *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science, Vol. VII (University of Minnesota Press, Mpls.) © 1975 University of Minnesota. Перевод выполнен по кн.: *Putnam H.* *Mind, Language and Reality*, Philosophical Papers, Cambridge, Mass., 1975, Vol. 2, Гл. 12, с. 215—272.

¹ Сегодня уже невозможно перечислить всех исследователей, внесших свой вклад в эту область, но первыми, конечно же, были Зелинг Харрис и Ноам Хомский.

² Обсуждение по этому вопросу см. в работе: *Putnam H.* The "innateness" hypothesis etc. // *Synthese*, 1967, Vol. 17, p. 12—22, *Homsky N.* *Problems of Knowledge and Freedom*. New York, 1971 (в частности, Гл. 1).

мантика находится в гораздо худшем состоянии, чем синтаксическая теория, состоит в том, что *донаучное* понятие, на которое опирается семантика, т. е. *донаучное* понятие *значения*, само находится в гораздо худшем состоянии, чем *донаучное* понятие синтаксиса. Как правило, в философии сомнения скептиков относительно понятия ничуть не лучше помогают прояснить или улучшить ситуацию, чем догматические заявления консервативно настроенных философов, что все замечательно в этом лучшем из возможных миров. Не раскрывают причину, почему *донаучное* понятие значения находится в плохом состоянии, и общие скептические или номиналистические доводы в пользу того, что значения не существуют. Итогом нашего рассмотрения, действительно, будет вывод о том, что значения не существуют в том смысле, в каком мы склонны думать, что они существуют. Однако электроны также не существуют в том смысле, в каком думал Бор. Есть немалая разница между этим моим утверждением и утверждением, что значения (или электроны) «не существуют».

Практически я буду говорить только о значении слов, а не о значении предложений, поскольку считаю, что наше понятие значения слова гораздо более несовершенно, чем понятие значения предложения. Но я кратко коснусь доводов таких философов, как Дональд Дэвидсон, настаивающих на том, что понятие значения слова *должно* быть вторичным, а исследование значения предложения — первичным. На мой взгляд, традиционные теории значения устраняют некоторые мифы (отмечу, что «значение» — единственная обсуждаемая в философии тема, в которой нет буквально ничего, кроме «теории»; буквально ничего, что можно было бы назвать, пусть даже в насмешку, «точкой зрения здравого смысла»), поэтому мне придется рассмотреть и попытаться разобраться с рядом тем, в отношении которых я считаю общепринятую точку зрения ложной. Читатель окажет мне максимальную помощь в решении этой задачи, если любезно согласится допустить, что *ничего* не ясно заранее.

ЗНАЧЕНИЕ И ЭКСТЕНСИОНАЛ

По крайней мере со средних веков авторы, пишущие по теории значения, стремились выявить двойственность в обыденном понятии значения и ввели пару терминов: экстенционал и интенционал, или *Sinn* и *Bedeutung* и т. п., в целях преодоления этой двойственности. *Экстенционал* термина, в обычном логическом употреблении, — это просто множество вещей, относительно которых данный термин истинен. Так, термин «заяц», в своем самом обычном смысле, исти-

нен относительно всех зайцев и только зайцев, поэтому именно множество зайцев и является экстенсионалом термина «заяц». Однако даже это понятие, *наименее* проблематичное в этом туманном предмете исследования, сталкивается с трудностями. Если не считать проблем, унаследованных экстенсионалом от понятия *истины*, приведенный пример термина «заяц», в его самом обычном смысле, обнаруживает одну из таких трудностей: строго говоря, он является не термином, а упорядоченной парой, состоящей из термина и «смысла» (или способа употребления (*occasion of use*), или чего-то такого, что позволяет отличить термин в одном смысле от того же самого термина, употребленного в другом смысле) и имеющей экстенсионал. Другая проблема такова: в математическом смысле «множество» — это объект типа «да—нет»: любая данная вещь или определенно принадлежит к S , или определенно не принадлежит к S , если S — множество. Однако слова естественного языка обычно не относятся к объектам типа «да—нет»: конечно, существуют вещи, относительно которых описание «дерево» определенно истинно, и вещи, относительно которых описание «дерево» определенно ложно, но существует и множество промежуточных случаев. Более того, граница между ясными и промежуточными случаями сама не является четкой. Таким образом, идеализация, связанная с понятием *экстенсионал* и предполагающая, что существует такая вещь, как множество объектов, относительно которых термин «дерево» истинен, по сути является весьма сильной.

Недавно ряд математиков исследовали понятие *нечеткого множества* (*buzzy set*), т. е. объекта, к которому другие вещи принадлежат или не принадлежат не с определенностью «да—нет», а с некоторой вероятностью или в некоторой степени. Если бы кто-нибудь захотел формализовать понятие экстенсионал применительно к терминам естественного языка, то ему потребовались бы «нечеткие множества» или что-то в этом роде, а не множества в их классическом смысле.

Проблему, связанную с тем, что слово имеет более одного смысла, обычно решают таким образом, что каждый смысл трактуется как отдельное слово (или, скорее, слово трактуется так, как если бы оно обладало невидимыми индексами: «заяц₁» — животное определенного вида; «заяц₂» — трус; и как если бы «заяц₁», «заяц₂» и т. д. были совершенно разными словами). Это, опять-таки, включает две (или по меньшей мере две) довольно сильные идеализации: предположение, что слова обладают множеством дискретных смыслов, и предположение, что полный набор смыслов установлен раз и

навсегда. Поль Зифф недавно показал, в какой степени оба эти предположения искажают реальное положение дел в естественном языке ³. Тем не менее, в настоящей статье мы будем исходить из этих идеализаций.

Теперь рассмотрим сложные термины «существо, имеющее сердце» и «существо, имеющее почки». Если допустить, что каждое существо, имеющее сердце, имеет и почки, и наоборот, то экстенсионалы этих двух терминов точно совпадают. Однако очевидно, что эти термины различаются по значению. Если предположить, что имеется смысл «значения», когда значение = экстенсионалу, то должен быть еще один смысл «значения», когда значением термина является не экстенсионал, а нечто иное, скажем «понятие», связанное с этим термином. Назовем это «нечто иное» *интенсионалом* термина. Понятие существа, имеющего сердце, определенно отличается от понятия существа, имеющего почки. Стало быть, два термина имеют различные интенсионалы. Когда мы говорим, что они имеют разные «значения», значение = интенсионалу.

ИНТЕНСИОНАЛ И ЭКСТЕНСИОНАЛ

Изложение, аналогичное предыдущему параграфу, можно встретить в любом месте, где стандартным образом разъясняются понятия «интенсионал» и «экстенсионал». Но это изложение вовсе не является удовлетворительным. Ответ на вопрос, почему оно не удовлетворительно, составляет, в некотором смысле, основное содержание настоящей статьи. Однако несколько моментов можно отметить с самого начала: прежде всего, какие есть свидетельства тому, что «экстенсионал» — это смысл слова «значение»? Каноническое объяснение понятий «интенсионал» и «экстенсионал» выглядит примерно так: «в одном смысле “значение” означает *экстенсионал*, а в другом смысле “значение” означает *значение*». Но дело в том, что, если понятие «экстенсионал» определено относительно точно на основе фундаментального логического понятие *истины* (и с использованием указанных выше сильных идеализаций), то понятие интенционала определено не точнее неясного (и, как мы увидим далее, вводящего в заблуждение) понятия «понятие». Это, как если бы кто-то объяснял понятие «вероятность» следующим образом: «в одном смысле “вероятность” означает частоту, а в другом смысле это по-

³ См. обсуждение этого вопроса в кн.: Ziff P. Understanding Understanding. N. Y., 1972 (в частности, Гл. VIII).

нятие означает *предрасположенность* (propensity). «Вероятность» *никогда* не означает «частоту», а «предрасположенность», по меньшей мере, столь же неясна, как и «вероятность».

Какой бы неясной ни была традиционная доктрина, утверждающая двойственность понятия «значения», обладающего экстенсионалом и интенсионалом, из нее вытекают некоторые характерные следствия. Большинство традиционных философов рассматривали понятия как нечто *ментальное*. Таким образом, доктрина, трактующая значение термина (т. е. значение «в смысле интенционала») как понятие, имела то следствие, что значения представляют собой ментальные сущности. Однако Фреге, а позднее Карнап и его последователи восстали против такого, как они говорили, «психологизма». Сознвая, что значения — это *общественное* достояние и что *одно и то же* значение может быть «усвоено» многими людьми и в разные периоды времени, они отождествляли понятия (а, следовательно, «интенционалы» или значения) не с ментальными, а с абстрактными сущностями. Однако «усвоение» этих абстрактных сущностей оставалось индивидуальным психологическим актом. Никто из этих философов не сомневался в том, что понимание слова (знание его интенционала) связано с пребыванием в определенном психологическом состоянии (примерно в том же смысле, в каком знание, как умножать в уме числа, связано с пребыванием в определенном, очень сложном психологическом состоянии).

Во-вторых, известный пример с двумя терминами «существо, имеющее почки» и «существо, имеющее сердце», показывает, что два термина могут иметь один и тот же экстенсионал и, тем не менее, различаться по интенсионалу. Но считалось очевидным, что обратное невозможно: два термина не могут различаться по экстенсионалу и иметь один и тот же интенсионал. Интересно, что никогда не предлагалось доказательства этой невозможности. Вероятно, в этом нашла отражение традиция античных и средневековых философов, полагавших, что понятие, соответствующее термину, представляет собой конъюнкцию предикатов, а, следовательно, это понятие должно *всегда* предоставлять необходимое и достаточное условие для вхождения объекта в экстенсионал термина ⁴. Для филосо-

⁴ Эта традиция выросла из анализа определенного термина — а именно, термина «Бог», который вызвал целую дискуссию в средневековой философии; считалось, что термин «Бог» определяется через конъюнкцию таких терминов, как «Благой», «Всемогущий», «Всезнающий» и т. п., т. е. через конъюнкцию так называемых «Совершенств». Однако возникла проблема в связи с тем, что Бога считали Единым, а Единство исключало *какую*

фов, которые, подобно Карнапу, приняли верификационную теорию значения, понятие, соответствующее термину, выступает (в идеальном случае, когда термин обладает «полным значением») *критерием* принадлежности к экстенсионалу (критерием, понимаемым не просто как «необходимое и достаточное условие», а в строгом смысле как *способ распознавания* того, входит ли данный предмет в экстенсионал или нет). Поэтому эти философы-позитивисты с радостью восприняли традиционную точку зрения в этом вопросе. Итак, в основе теории значения лежат два не вызывавших сомнения допущения:

(1) Знание значения термина связано с пребыванием в определенном психологическом состоянии (где «психологическое состояние» понимается в том смысле, в каком «психологическими» являются состояния памяти и психологические предрасположенности; никто, конечно, не считал, что знание значения слова — это длящееся состояние осознания).

(2) Значение термина (понятое как «интенсионал») определяет его экстенсионал (в том смысле, что из тождества интенционалов вытекает тождество экстенционалов).

Я постараюсь показать, что никакое понятие, не говоря уже о понятии значения, не удовлетворяет этим двум вместе взятым допущениям. Традиционное понятие значения имеет в своем основании ложную теорию.

бы то ни было сложность Его сущности. Таким образом, «Бог» определялся через конъюнкцию терминов, но Бог (без кавычек) не мог быть ни логическим произведением свойств, ни уникальной вещью, являющей собой логическое произведение двух и более *отдельных* свойств, поскольку даже этот очень абстрактный вид «сложности» считался несовместимым с совершенством Единого. Над этим теологическим парадоксом веками бились еврейские, арабские и христианские теологи (достаточно упомянуть доктрину отрицательных определений Бога, разрабатываемую Маймонидам и Аквинатом). Примечательно, что такие теории, как концептуализм и номинализм — представляющие интерес и сегодня, — сначала были предложены в качестве решения проблемы предикации применительно к Богу. Не менее примечательно и то, что излюбленная для всей этой теологии модель определения (модель в виде конъюнкции-всех-свойств) сохранилась, по крайней мере через свои следствия, в философии языка вплоть до настоящего времени.

**«ПСИХОЛОГИЧЕСКОЕ СОСТОЯНИЕ»
И МЕТОДОЛОГИЧЕСКИЙ СОЛИПСИЗМ**

Для доказательства этого нам нужно сначала прояснить традиционное понятие психологического состояния. В одной трактовке, это состояние представляет собой двухместный предикат, аргументами которого являются индивид и время. В этом смысле «*быть ростом в пять футов*», «*испытывать боль*», «*знать алфавит*» и даже «*быть за тысячу миль от Парижа*» — это все состояния. (Отметим, что *время* обычно остается имплицитным или «контекстуальным»; полная форма атомарного предложения для этих предикатов будет такова: «*x ростом в пять футов во время t*», «*x испытывает боль во время t*» и т. д.) Однако в науке термин состояния обычно применяется только к свойствам, определяемым в терминах фундаментальных с точки зрения данной науки параметров индивида. Так, быть ростом в пять футов — это состояние (с точки зрения физики); испытывать боль — тоже состояние (по крайней мере, с точки зрения менталистской психологии); знать алфавит — можно, хотя и с некоторой натяжкой, считать состоянием (с точки зрения когнитивной психологии); но быть за тысячу миль от Парижа будет неестественно назвать *состоянием*. В другой трактовке психологическое состояние — это просто состояние, изучаемое и описываемое психологией. В этом смысле может быть тривиальным утверждение, что *знание значения слова «вода»* является «психологическим состоянием» (с точки зрения когнитивной психологии). Но в указанном выше допущении (1) говорится о психологическом состоянии не в этом смысле.

Когда традиционные философы рассуждают о психологических состояниях (или «ментальных» состояниях), они исходят из допущения, которое можно назвать допущением методологического солипсизма. Согласно этому допущению, никакое психологическое состояние, в собственном смысле слова, не предполагает существования индивидов, помимо того индивида, кому это состояние приписывается. (Фактически, это допущение гласило, что никакое психологическое состояние не предполагает существования даже *тела* индивида: если *P* — психологическое состояние в собственном смысле слова, то должно быть логически возможным пребывание «бестелесного разума» в этом состоянии *P*). Это допущение довольно явным образом представлено у Декарта, но оно в скрытом виде присутствует и во всей традиции философской психологии. Принимая это допущение, мы неизбежно подписываемся под программой *огра-*

ничения психологии и сознательно устанавливаем рамки, в которых характер и область психологического исследования должны согласовываться с менталистскими предрассудками, а порой и с идеалистическим истолкованием знания и мира. Однако часто не замечают, сколь серьезные ограничения накладывает эта программа. Например, если принять методологический солипсизм, то потребуются переинтерпретировать такие распространенные и заурядные психологические состояния как состояние *«ревности»*. Из утверждения *«x ревнует y»*, в его обычном употреблении, следует, что существует *y*, а из утверждения *«x ревнует y к z»* следует, что существуют и *y*, и *z* (не говоря уже об *x*). Стало быть, с точки зрения методологического солипсизма *«ревновать»* и *«ревновать кого-то к кому-то»* не являются допустимыми психологическими состояниями. (Мы будем называть их «психологическими состояниями в широком смысле», а состояния, допустимые с точки зрения методологического солипсизма, — «психологическими состояниями в узком смысле»). Методологический солипсизм потребовал бы так переинтерпретировать *«ревность»*, чтобы можно было говорить, что я ревную свои собственные галлюцинации или плоды собственного воображения и т. п. Только если допустить, что психологические состояния в узком смысле обладают определенной степенью каузальной замкнутости (и, поэтому, ограничившись психологическими состояниями в узком смысле, мы сможем сформулировать психологические законы), — только в этом случае подобная переинтерпретация или принятие методологического солипсизма будут иметь смысл. Однако три века неудач менталистской психологии, на мой взгляд, со всей очевидностью свидетельствуют против такого подхода.

Как бы то ни было, мы можем теперь точнее сформулировать тезисы, выдвинутые в конце предыдущего раздела. Пусть *A* и *B* — любые два термина, различающиеся по экстенционалу. Согласно допущению (2) они должны различаться по значению (понятому как «интенционал»). Согласно допущению (1) *знание значения A* и *знание значения B* являются психологическими состояниями в узком смысле, поскольку именно так мы трактуем допущение (1). Однако эти психологические состояния должны в той же мере определять экстенционалы терминов *A* и *B*, в какой их определяют значения («интенционалы»).

Чтобы доказать это, попробуем допустить обратное. Безусловно, не может быть двух таких терминов *A* и *B*, чтобы *знание значения термина A* представляло собой то же самое состояние, что и *знание значения термина B*, хотя *A* и *B* имели бы различные экстенциона-

лы. *Знание значения термина A — это не просто «усвоение интенционала» термина A , что бы это ни означало; это также знание того, что «усвоенный» «интенционал» является интенционалом термина A . (Стало быть, тот, кто знает значение слова «колесо», по-видимому, знает и «интенционал» его немецкого синонима *Rad*; но если он не знает, что данный «интенционал» является интенционалом слова *Rad*, об этом человеке не скажут, что он «знает значение *Rad*».) Если A и B — разные термины, то *знание значения термина A* является другим состоянием по сравнению со *знанием значения термина B* , независимо от того, совпадают или различаются по значению сами термины A и B . Таким же способом можно показать, что если I_1 и I_2 — разные интенционалы, а A — термин, то *знание того, что I_1 является значением A* представляет собой другое психологическое состояние по сравнению со *знанием того, что I_2 является значением A* . Таким образом, не может быть двух разных возможных миров L_1 и L_2 таких, что, скажем, Оскар находится в *одинаковом во всех отношениях психологическом состоянии* (в узком смысле) и в L_1 , и в L_2 , и при этом в L_1 Оскар понимает термин A в значении I_1 , а в L_2 он понимает A в значении I_2 . (Будь это так, психологическим состоянием Оскара в L_1 было бы *знание того, что I_1 является значением A* , а психологическим состоянием Оскара в L_2 было бы *знание того, что I_2 является значением A* , а это два разных и даже — если допустить, что A имеет только одно значение для Оскара в каждом мире, — несовместимых психологических состояния в узком смысле.)*

Короче говоря, если S — психологическое состояние рассматриваемого нами вида, т. е. если оно представляет собой *знание того, что I является значением A* , где I — «интенционал», а A — термин, то *во всех возможных мирах, в которых говорящий находится в психологическом состоянии S , «работают» одни и те же необходимые и достаточные условия вхождения в экстенционал термина A , поскольку состояние S определяет интенционал I , а, согласно допущению (2), интенционал выражает необходимые и достаточные условия вхождения в экстенционал.*

Если наша интерпретация традиционного учения об интенционале и экстенционале справедлива в отношении Фреге и Карнапа, то применительно к теории значения антитеза «психологизм/платонизм» оказывается чем-то вроде бури в стакане воды. (Несомненно, эта антитеза образует очень важную проблему, когда речь идет об общей философии математики.) Даже если значения, как считают Фреге и Карнап, представляют собой «платоновские», а не «мен-

тальные», сущности, «усвоение» этих сущностей, по-видимому, является психологическим состоянием (в узком смысле). Более того, психологическое состояние определяет уникальным образом «платоновскую» сущность. Поэтому видеть ли в «значении» «платоновскую» сущность или психологическое состояние — это уже вопрос конвенции. Выбор психологического состояния в качестве значения едва ли будет иметь столь пугавшее Фреге следствие, что значения утратят свой общественный характер. Ибо и психологические состояния являются «общественными» в том смысле, что разные люди (и даже в разные эпохи) могут находиться в *одном и том же* психологическом состоянии. Фактически, аргумент Фреге против психологизма направлен только против отождествления понятий с конкретными сущностями, а не с ментальными сущностями вообще.

Из «общественного» характера психологических состояний следует, в частности, то, что, если Оскар и Элмер понимают слово *A* по-разному, то они должны находиться в разных психологических состояниях. Ибо *состояние знания того, что интенционалом термина A является, скажем, I — это одно и то же состояние*, независимо от того, находится в нем Оскар или Элмер. Поэтому двое говорящих не могут находиться в одинаковом — во всех отношениях — психологическом состоянии и понимать термин *A* по-разному; психологическое состояние говорящего определяет интенционал (*a*, следовательно, согласно допущению (2), и экстенционал) термина *A*.

Именно это последнее следствие из допущений (1) и (2), вместе взятых, мы и считаем ложным. Мы утверждаем, что два человека могут находиться в одном и том же психологическом состоянии (в узком смысле), даже если экстенционал термина *A* в идиолекте одного отличается от экстенционала термина *A* в идиолекте другого. Экстенционал *не* определяется психологическим состоянием.

Более подробно мы покажем это в последующих разделах. Если это верно, то перед каждым, кто хочет спасти хотя бы одно из традиционных допущений, открываются два пути: или отбросить идею, что психологическое состояние (в узком смысле) определяет *интенционал*, или отбросить идею, что интенционал определяет экстенционал. Мы рассмотрим эти альтернативы позже.

НАХОДЯТСЯ ЛИ ЗНАЧЕНИЯ В ГОЛОВЕ?

Мы покажем, что психологическое состояние не определяет экстенционал, прибегнув к небольшой научно-фантастической исто-

рии. Предположим, что где-то в галактике существует планета, которую мы будем называть Двойником Земли. Двойник Земли очень похож на Землю вплоть до того, что люди на Двойнике Земли говорят *по-английски*. Фактически, если оставить в стороне различия, о которых мы скажем особо, читатель может считать Двойник Земли *точной* копией Земли. При желании он может даже допустить, что у него есть *Doppelgänger*⁶ на Двойнике Земли; хотя для моего сюжета это не важно.

Хотя некоторые люди на Двойнике Земли (скажем, и те, кто называет себя «американцами», и те, кто называет себя «канадцами», и те, кто называют себя «англичанами» и т. п.) говорят *по-английски*, нет ничего удивительного в том, что есть некоторые небольшие различия между английскими диалектами на Двойнике Земли и литературным английским. Эти различия обусловлены некоторыми особенностями Двойника Земли.

Одна из этих особенностей Двойника Земли состоит в том, что здесь «водой» называется не жидкость H_2O , а другая жидкость с очень длинной и сложной химической формулой. Для краткости я запишу эту формулу как XYZ. Предполагается, что при нормальной температуре и давлении XYZ неотличима от воды. В частности, она имеет вкус воды и утоляет жажду как вода. Предполагается также, что океаны, озера и моря на Двойнике Земли заполнены не водой, а жидкостью XYZ, и в виде дождя там выпадает не вода, а XYZ и т. п.

Если когда-нибудь на Двойник Земли прилетит космический корабль с Земли, то люди с этого корабля сначала решат, что слово «вода» на Земле и на Двойнике Земли имеет одно и то же значение. Это предположение будет исправлено, когда обнаружится, что «вода» на Двойнике Земли — это XYZ, и сообщение землян с корабля будет примерно таким:

«На Двойнике Земли слово “вода” означает XYZ».

(Кстати, именно такого рода употребление слова «означает» служит основанием для утверждения, что экстенционал — это один из смыслов «значения». Однако заметьте, что, хотя слово «означает» в этом примере действительно означает нечто вроде *иметь экстенционал*, мы, видимо, не сказали бы, что

«На Двойнике Земли значением слова “вода” является XYZ».

⁶ Двойник (нем. яз.) — прим. перев.

если бы каждому взрослому жителю на Двойнике Земли не был известен тот факт, что «вода есть XYZ». Мы сможем объяснить это с помощью предлагаемой ниже теории значения; сейчас же отметим только, что хотя глагол «означать» иногда означает «иметь экстенционал», образованное от него существительное «значение» *никогда* не означает «экстенционал».)

Соответственно, если когда-нибудь на Землю прилетит космический корабль с Двойника Земли, то космонавты сначала сочтут, что слово «вода» на Двойнике Земли и на Земле имеет одно и то же значение. Это предположение будет исправлено, когда обнаружится, что «вода» на Земле — это H_2O , и космонавты пошлют сообщение:

«На Земле ⁶ слово “вода” означает H_2O ».

Заметьте, что экстенционал термина «вода» не вызывает проблем. Это слово просто имеет (как мы говорим) два разных значения: на Двойнике Земли оно употребляется в смысле воды_{дз} — и поэтому то, что мы на Земле называем «водой», — не вода для жителей Двойника Земли, а на Земле оно употребляется в смысле воды_з, и поэтому то, что люди с Двойника Земли называют «водой», — не вода для нас. Экстенционалом «воды» в смысле воды_з является множество всех объектов, состоящих из молекул H_2O , или что-то в этом роде; экстенционалом воды в смысле воды_{дз} является множество всех объектов, состоящих из молекул XYZ, или что-то в этом роде.

Теперь отмотаем время назад в 1750 г. Тогда химия не была развитой наукой ни на Земле, ни на Двойнике Земли. Житель Земли не знал, что вода состоит из водорода и кислорода, а житель Двойника Земли не знал, что «вода» состоит из молекул XYZ. Пусть Оскар₁ будет типичным жителем Земли, а Оскар₂ — его двойником на Двойнике Земли. Можно считать, что Оскар₁ знает о воде то же самое, что Оскар₂ знает — о «воде». Если хотите, можно даже предположить, что Оскар₁ и Оскар₂ — точные подобию друг друга по внешности, чувствам, мыслям, внутренним монологам и т. д. И тем не менее, экстенционалом термина «вода» на Земле была та же самая жидкость H_2O как в 1750 г., так и в 1950 г.; а экстенционалом термина «вода» на Двойнике Земли была та же самая жидкость XYZ, как в 1750 г., так и в 1950 г., Оскар₁ и Оскар₂ понимали по-разному термин «вода» в 1750 г., *хотя они находились в одном и*

⁶ Или, скорее, они напишут: «На Двойнике Земли (название, данное Земле на Двойнике Земли — X. П.) слово “вода” означает H_2O ».

том же психологическом состоянии, и хотя при тогдашнем состоянии науки ученым потребовалось бы еще приблизительно 50 лет, чтобы установить, что они по-разному понимают термин «вода». Таким образом, экстенционал термина «вода» (а фактически и то, что в интуитивном, доаналитическом смысле называется «значением») не является сам по себе функцией от психологического состояния говорящего.

Здесь можно было бы возразить: почему мы должны считать, что термин «вода» в 1750 г. и в 1950 г. имел один и тот же экстенционал (и на Земле, и на Двойнике Земли)? С точки зрения логики термины естественных видов вроде «воды» представляют собой очень сложный предмет исследования, и предполагаемое ниже — лишь набросок ответа. Предположим, я указываю на стакан воды и говорю: «эта жидкость называется водой» (или «это называется водой», если маркер «жидкость» ясен из контекста). Мое «остенсивное определение» воды имеет следующую эмпирическую предпосылку: тот объем жидкости, на который я указываю, находится в определенном отношении тождества (скажем, x есть «та же самая жидкость», что и y) к большей части вещества, которое я и другие люди в моем языковом сообществе в других случаях называют водой. Если эта предпосылка ложна — скажем, потому что я, сам не зная того, указываю на стакан с джином, а не с водой, — тогда я не могу рассчитывать, что мое остенсивное определение будет принято. Таким образом, остенсивное определение содержит, так сказать, отменяемое (*defeasible*) необходимое и достаточное условие: необходимым и достаточным условием для того, чтобы быть водой, является наличие отношения «та же самая жидкость» к содержимому стакана; однако, это условие необходимо и достаточно только в том случае, если истинна эмпирическая предпосылка. Если она не истинна, то вступает в действие одно из так называемых «запасных» (ball-back) условий.

Здесь существенно то, что отношение «та же самая жидкость» является *теоретическим* отношением: чтобы установить, имеет ли место это отношение между какой-то жидкостью и *этой* жидкостью, может потребоваться какое угодно количество научных исследований. Более того, даже если в ходе научного исследования или в результате применения критерия «здравого смысла» и будет получен «определенный» ответ, его нельзя считать окончательным: будущее исследование может аннулировать даже самый «определенный» случай. Стало быть, тот факт, что люди в 1750 г. могли называть жидкость XYZ «водой», в то время как в 1800 г. или в 1850 г. другие

поколения называли «водой» жидкость H_2O , не означает, что для среднего человека за этот промежуток времени «значение» слова «вода» изменилось. И в 1750 г., и в 1850 г., и в 1950 г. можно было, скажем, указать на жидкость в озере Мичиган как на пример «воды». Изменилось то, что в 1750 г. мы ошибочно считали, что XYZ находится в отношении «та же самая жидкость» к жидкости в озере Мичиган, тогда как в 1800 г. или в 1850 г. мы уже знали, что это не так. (Я игнорирую тот факт, что в 1950 г. жидкость в озере Мичиган уже сомнительно называть водой.)

А теперь давайте несколько изменим нашу научно-фантастическую историю. Я не знаю, можно ли делать кастрюли и сковородки из молибдена; равно как я не знаю и того, легко ли отличить кастрюли и сковородки из молибдена от кастрюль и сковородок, сделанных из алюминия. (Я ничего этого не знаю, хотя усвоил слово «молибден».) Поэтому я буду считать, что отличить кастрюли и сковородки, сделанные из молибдена, от кастрюль и сковородок, сделанных из алюминия, может *только* специалист. (Подчеркну, что, насколько я могу судить, так могло бы быть, и *a fortiori* так могло бы быть, если учесть все, что мне известно благодаря «знанию значения» слов *алюминий* и *молибден*.) Теперь представим себе, что на Двойнике Земли молибден так же повсеместно распространен, как на Земле алюминий, а алюминий на Двойнике Земли столь же редок, как на Земле — молибден. В частности, мы будем считать, что «алюминиевые» сковородки и кастрюли на Двойнике Земли сделаны из молибдена. Наконец, будем считать, что слова «алюминий» и «молибден» *поменялись местами* на Двойнике Земли: «алюминий» — это название *молибдена*, а «молибден» — название *алюминия*.

Этот пример в чем-то похож на предыдущий. Если бы на Двойник Земли прилетел космический корабль с Земли, то земляне, возможно, и не заподозрили бы, что «алюминиевые» сковородки и кастрюли на Двойнике Земли сделаны не из алюминия, тем более, что жители Двойника Земли говорили бы, что они сделаны из него. Однако между этими двумя примерами имеется одно важное различие. Металлург с Земли мог бы определить, что «алюминий» — это молибден, а металлург с Двойника Земли мог бы легко установить, что алюминий — это «молибден». (В этом предложении устрашающие кавычки указывают на то, что слово употреблено в значении, которое оно имеет на Двойнике Земли.) В то время, как в 1750 г. никто ни на Земле, ни на Двойнике Земли не мог бы отличить воду от «воды», алюминий с «алюминием» спутали бы только некоторые люди из этих языковых сообществ.

Данный пример иллюстрирует ту же мысль, что и предыдущий. Если Оскар₁ и Оскар₂ представляют собой стандартных носителей языка на Земле и Двойнике Земли, соответственно, и ни один из них не имеет особых познаний в химии или металлургии, то может и не быть никаких отличий в их психологическом состоянии, когда они употребляют слово «алюминий»; тем не менее, мы должны констатировать, что слово «алюминий» имеет экстенционал *алюминия* в идиолекте Оскара₁ и экстенционал *молибдена* в идиолекте Оскара₂. (Приходится также констатировать, что Оскар₁ и Оскар₂ подразумевают под словом «алюминий» разные вещи, и что «алюминий» имеет на Земле иное значение, чем на Двойнике Земли и т. д.) Вновь мы убеждаемся, что психологическое состояние говорящего *не* определяет экстенционал слова (или его «значение», если воспользоваться словом из «доаналитического» лексикона).

Прежде чем продолжить рассмотрение этого примера, позвольте мне предложить пример *не* из научной фантастики. Предположим, что вы, как и я, не можете отличить вяз от бука. Все же мы говорим, что экстенционал слова «вяз» в моем идиолекте такой же, как и в идиолекте любого другого человека, а именно — все множество вязов; и что все множество буков образует экстенционал слова «бук» в *обоих* наших идиолектах. Таким образом, экстенционал слова «вяз» в моем идиолекте отличается от экстенционала слова «бук» в вашем идиолекте (как и должно быть). Неужели есть основания утверждать, что это различие в экстенционалах обусловлено различием в наших *понятиях*? Мое *понятие* о вязе в точности совпадает с моим *понятием* о буке (в чем я со стыдом признаюсь). (Кстати, это показывает, что отождествление значения «в смысле интенционала» с *понятием* нельзя считать корректным.) Если же кто-то героически продолжает утверждать, что различие между экстенционалом слова «вяз» и экстенционалом слова «бук» в моем идиолекте объясняется различием в моем психологическом состоянии, то мы всегда можем доказать его неправоту, приведя пример с «Двойником Земли» и допустив, что слова «вяз» и «бук» на Двойнике Земли поменялись местами (как слова «алюминий» и «молибден» в предыдущем примере). Более того, предположим, что у меня есть *Doppelgänger* на Двойнике Земли, молекула в молекулу «тождественный» мне (в том смысле, в каком могут быть «тождественны» два «галстука»). Если вы дуалист, то можете вдобавок предположить, что мой *Doppelgänger* продумывает те же самые мысли и в тех же самых словах, что и я, имеет те же самые ощущения и те же самые склонности, какие имею я, и т. д. Было бы нелепо думать, что его психо-

логическое состояние в чем-то отличается от моего: и тем не менее, он «имеет в виду» *бук*, когда говорит «вяз», а я «имею в виду» *вяз*, когда говорю *вяз*. Как тут ни крути, но «значения» не находятся в голове!

СОЦИОЛИНГВИСТИЧЕСКАЯ ГИПОТЕЗА

Два последние примера указывают на важный для языка факт, который, как это ни удивительно, никогда прежде, кажется, не отмечался: существует *разделение лингвистического труда*. Мы вряд ли могли бы пользоваться такими словами как «вяз» и «алюминий», если бы никто не умел распознавать среди деревьев вязы и среди металлов алюминий; но отнюдь не каждый, для кого это важно, должен сам уметь распознавать их. Возьмем другой пример — *золото*. Золото важно во многих отношениях: это драгоценный металл, который используется для изготовления денег, оно имеет символическую ценность (для большинства людей важно, чтобы их «золотое» обручальное кольцо *действительно* было из золота, а не просто *выглядело* золотым), и т. д. Наше сообщество можно считать своего рода «фабрикой»: на этой «фабрике» одни люди «заняты» *ношением золотых обручальных колец*, другие «заняты» их *продажей*, а третьи «заняты» тем, что *устанавливают, является ли тот или иной предмет действительно золотым*. Совершенно обязательно и нецелесообразно, чтобы каждый, кто носит золотое кольцо (или золотые заколки и т. п.) или обсуждает «золотой стандарт» и т. п., занимался бы *куплей-продажей* золота. Нет нужды и в том, чтобы каждый, кто покупает и продает золото, был способен установить, является ли вещь действительно золотой, если в обществе не практикуется такая форма обмана, как продажа фальшивого золота, а в случае сомнения каждый всегда может проконсультироваться у эксперта. И, *конечно же*, необязательно и нецелесообразно, чтобы каждый, кто покупает или носит золотые вещи, был способен с достоверностью установить, являются ли они действительно золотыми или нет.

Все эти факты просто иллюстрируют обычное разделение труда (в широком смысле). Но они предполагают и разделение лингвистического труда: каждый, для кого по какой-либо причине важно *золото*, должен овладеть словом «золото»; но он вовсе не обязан владеть *методом распознавания* золота. В этом он может положиться на особый подкласс носителей языка. Признаки, которые обычно связываются с общим именем, т. е. необходимые и достаточные усло-

вия для включения объекта в экстенционал имени, способы распознавания его вхождения в экстенционал («критерии») и т. д., — все это известно лингвистическому сообществу, *рассматриваемому как коллективное тело*; но этому коллективному телу свойственно разделение «труда» на виды, связанные со знанием и использованием различных аспектов «значения» слова (например, «золота»).

Разумеется, это разделение лингвистического труда предполагает разделение *нелингвистического* труда и основывается на нем. Если считать, что слово «золото» может входить только в словарь тех людей, которые умеют отличать золото от других металлов, то по отношению к этому подклассу людей слово «золото» будет в том же положении, что и слово «вода» в 1750 г., а остальные люди вообще не будут владеть этим словом. Некоторые слова не предполагают разделения лингвистического труда: например, «стул». Но с ростом общественного разделения труда и развитием науки увеличивается число слов, требующих разделения лингвистического труда. Слово «вода», к примеру, никак не обнаруживало этого разделения труда до возникновения химии. Совершенно очевидно, что сегодня каждый носитель языка должен уметь распознавать воду (в нормальных условиях безошибочно); возможно даже, каждый взрослый носитель языка знает необходимое и достаточное для идентификации воды условие: «вода есть H_2O »; но лишь немногие взрослые люди смогли бы отличить воду от жидкостей, которые внешне ее напоминают. В случае сомнения другие люди полагались бы на суждение этих «экспертов». Таким образом, способом распознавания, которым владеют эти «эксперты», благодаря им, владеет и все лингвистическое сообщество, хотя отнюдь не всякий индивидуальный представитель этого сообщества умеет его применять; в результате самые специальные сведения о воде могут стать частью *общественного* значения этого слова, хотя они и неизвестны большинству людей, употребляющих его.

Думается, что разделение лингвистического труда представляет важное явление для социолингвистического исследования. В этой связи я хотел бы предложить следующую гипотезу:

ГИПОТЕЗА ОБ УНИВЕРСАЛЬНОСТИ РАЗДЕЛЕНИЯ ЛИНГВИСТИЧЕСКОГО ТРУДА: В каждом лингвистическом сообществе имеет место описанное выше разделение лингвистического труда, то есть в любом таком сообществе используется хотя бы несколько терминов, с которыми связаны «критерии», известные только одной подгруппе носителей языка, употребление этих терминов остальными людьми

предполагает их сотрудничество с носителями языка из соответствующих подгрупп.

Было бы особенно интересно установить, составляют ли первобытные народы исключение из этой гипотезы (это означало бы, что разделение лингвистического труда представляет собой продукт социальной эволюции), или даже они обнаруживают это разделение. В последнем случае можно было бы выдвинуть предположение, что разделение труда, включая лингвистический труд, образует фундаментальную особенность человеческого рода.

Нетрудно понять, как это явление объясняет некоторые из приведенных выше примеров, которые показывают несостоятельность допущений (1) и (2). Когда употребление термина предполагает разделение лингвистического труда, «средний» носитель языка не располагает ничем, что фиксировало бы экстенционал такого термина. В частности, его индивидуальное психологическое состояние *определенно* не фиксирует экстенционал этого термина; только социолингвистическое состояние всего лингвистического сообщества, к которому принадлежит данный носитель языка, определяет экстенционал этого термина.

Подводя итог сказанному, можно отметить, что в мире существует два рода орудий: есть такие орудия, как молоток и отвертка, которыми может пользоваться один человек, и есть такие орудия, как пароход, для использования которых требуется совместная деятельность многих людей. До сих пор в словах главным образом видели орудия первого рода.

ИНДЕКСАЛЬНОСТЬ И ЖЕСТКОСТЬ ⁷

Первый из наших научно-фантастических примеров — о «воде» на Земле и на Двойнике Земли в 1750 г. — не предполагает разделения лингвистического труда или, по крайней мере, предполагает его не так, как примеры с «алюминием» и «вязом». В 1750 г. ни на Земле, ни на Двойнике Земли не было (во всяком случае, согласно нашей истории) никаких «экспертов» по воде. (Впрочем, пример можно переформулировать так, чтобы он предполагал разделение труда во *временной перспективе*. Но я не буду развивать такой

⁷ Содержание данного раздела образуют лекции, прочитанные мной в Вашингтонском Университете (Летний институт философии) в 1968 г., и лекция, прочитанная в Университете шт. Миннесота.

подход.) Вместе с тем, этот пример предполагает вещи, имеющие принципиально важное значение для теории референции, равно как для теории необходимой истины, о которой и пойдет речь.

Мы можем сообщить кому-либо, что означают такие термины естественных видов, как «вода», «тигр» или «лимон», двумя довольно очевидными способами. Во-первых, можно предложить так называемое остенсивное определение: «это (эта жидкость) — вода», «это (это животное) — тигр», «это (этот фрукт) — лимон»; где скобки означают, что «маркеры» *жидкость, животное, фрукт* могут указываться явным образом или могут предполагаться неявно. Во-вторых, можно дать *дескрипцию*. В последнем случае предлагаемая дескрипция обычно содержит один или несколько маркеров вместе со *стереотипом*, т. е. стандартизированным описанием типичных, «нормальных», или, во всяком случае, стереотипных признаков данного вида. Главные признаки, включенные в стереотип, как правило, выступают *критериями*, т. е. в обычных ситуациях они являются способом установления принадлежности той или иной вещи к данному виду, или, по крайней мере, составляют необходимые условия (или вероятностные необходимые условия) ее принадлежности к данному виду. Не все критерии, используемые лингвистическим сообществом как неким коллективным телом, входят в стереотип; в некоторых случаях стереотипы могут быть довольно слабыми. Так, для меня (если только я не являюсь нетипичным носителем языка) стереотип вяза — это стереотип обычного листопадного дерева. Эти признаки действительно образуют необходимое условие принадлежности к данному виду (я употребляю слово «необходимый» в очень широком смысле; я не считаю утверждение «вяза — это листопадные деревья» *аналитическим*), но их явно не хватает для распознавания вязов. С другой стороны, стереотип тигра позволяет распознавать тигров (если только они не являются альбиносами или если не имеет места еще какой-нибудь нетипичный случай), и стереотип лимона обычно позволяет распознавать лимоны. В крайнем случае стереотип может содержать один *лишь* маркер: так, стереотип молибдена может указывать только на то, что молибден — это *металл*. Рассмотрим оба эти способа введения термина в словарь каково-либо человека.

Допустим, в целях обучения кого-либо слову «вода» я указываю на стакан с жидкостью и говорю «это — вода». Мы уже описали некоторые эмпирические предпосылки этого акта и указали, как этот вид разъяснения значения может быть аннулирован. Теперь мы постараемся пояснить дальше, как это следует понимать.

В дальнейшем мы будем использовать понятие «возможного мира» в качестве исходного, поскольку считаем, что в некоторых отношениях это понятие оправдано и является важным с научной точки зрения, хотя оно нуждается в уточнении. Допустим также, что по крайней мере в некоторых случаях можно говорить об одном и том же человеке, что он существует более, чем в одном из возможных миров⁸. Наше изложение здесь в значительной степени опирается на работы Сола Крипке, хотя выводы были получены нами самостоятельно.

Обозначим как W_1 и W_2 два возможных мира, в которых существуют я, существует этот стакан и в которых, объясняя значение слова «вода», я указываю на этот стакан и говорю: «это вода». (Мы не предполагаем, что жидкость в стакане является одинаковой в обоих мирах.) Допустим, что в W_1 стакан наполнен H_2O , а в W_2 он наполнен XYZ . Предположим также, что W_1 — это *действительный* мир и что жидкость XYZ в мире W_2 обычно называется «водой» (так что отношение между носителями языка в W_1 и носителями языка в W_2 точно такое же, как отношение между носителями языка на Земле и носителями языка на Двойнике Земли). Тогда возможны две теории относительно значения слова «вода».

(1) Можно считать, что «вода» — это *соотносимый с разными мирами* (world-relative), но *постоянный* по значению термин (т. е. это слово имеет *постоянное относительное значение* (constant relative meaning)). Согласно этой теории слово «вода» *означает одно и то же* в W_1 и в W_2 ; просто в W_1 вода — это H_2O , а в W_2 вода — это XYZ .

(2) Можно считать, что вода — это H_2O во всех мирах (жидкость, называемая «водой» в W_2 , не является водой), но слово «вода» не имеет одного и того же значения в W_1 и в W_2 .

Если сказанное выше о Двойнике Земли верно, то очевидно, что верна теория (2). Когда я говорю: «это (эта жидкость) — вода», слово «это» употреблено мной, так сказать, в значении *de re*, т. е. убедительность моего разъяснения значения связана с тем, что словом «вода» обозначается все то, что находится в определенном отношении эквивалентности (в отношении, которое выше было названо «та же самая жидкость») к тому объему жидкости, на который указывает слово «это» в *действительном мире*.

⁸ На самом деле, это допущение не требуется для последующего доказательства: что нам *нужно*, так это возможность существования одного и того же *естественного* вида более, чем в одном возможном мире.

В символическом виде различие между двумя теориями может быть представлено как различие в «области действия». Согласно теории (1), истинно, что:

(1') (Для каждого мира W) (Для каждого x в W) (x есть вода $\equiv x$ находится в отношении «та же самая жидкость» к сущности, обозначенной словом «это» в мире W);

тогда как по теории (2):

(2') (Для каждого мира W) (Для каждого x в W) (x есть вода $\equiv x$ находится в отношении «та же самая жидкость» к сущности, обозначенной словом «это» в *действительном мире* W_1).

(Я называю это различием в «области действия», потому что в (1') «сущность, обозначенная словом “это”», находится в области действия квантора «Для каждого мира W », что явствует из уточняющего выражения «в мире W », тогда как в (2') «сущность, обозначенная словом “это”» означает «сущность, обозначенную словом “это” в *действительном мире*», то есть имеет референцию, *независимую* от связанной переменной « W ».)

Крипке называет десигнатор «жестким» (в данном предложении), если (в этом предложении) он обозначает один и тот же индивидуальный объект во всех возможных мирах, в которых этот десигнатор что-либо обозначает. Если мы распространим понятие жесткости и на имена веществ, то мы выразим обе теории — Крипке и мою, — сказав, что термин «вода» является *жестким десигнатором*.

Жесткость термина «вода» объясняется тем фактом, что, давая остенсивное определение «это (эта жидкость) — вода», я имею в виду (2'), а не (1').

Можно также сказать, вслед за Крипке, что в остенсивном определении «это (эта жидкость) — вода» указательное местоимение «это» является *жестким десигнатором*.

Крипке первым заметил, что эта теория значения (или «употребления» (название несущественно)) для слова «вода» (как и для других терминов естественных видов) имеет ошеломляющие последствия для трактовки необходимой истины.

Чтобы объяснить это, я введу понятие *кросс-мирового отношения* (*cross-world relation*). Будем называть двучленное отношение R кросс-мировым, если его экстенционалом является множество упорядоченных пар индивидуальных объектов, *не все из которых*

существуют в одном и том же возможном мире. Например, нетрудно истолковать отношение «того же роста, что и» как кросс-мировое отношение: это следует понимать так, что если x — индивид в мире W_1 ростом в пять футов, а y — индивид в мире W_2 ростом в пять футов, то упорядоченная пара x, y принадлежит к экстенсionalу отношения «того же роста, что и». (Поскольку индивид может иметь разный рост в разных возможных мирах, в которых он существует, то, строго говоря, не упорядоченная пара x, y является элементом экстенсionalа отношения «того же роста, что и», а упорядоченная пара x -в-мире- W_1, y -в-мире- W_2 .)

Аналогичным образом, мы можем трактовать отношение «та же самая жидкость» как кросс-мировое отношение, понимая его так, что жидкость в мире W_1 , имеющая те же существенные физические свойства (в W_1), какими обладает жидкость в W_2 , находится к этой последней в отношении «та же самая жидкость».

Итак, обобщая сказанное, нашу теорию можно сформулировать так: вещество x в любом произвольно взятом возможном мире, является *водой*, если и только если оно находится в отношении «та же самая жидкость» (понимаемом как кросс-мировое отношение) к тому веществу, которое мы называем «водой» в *действительном* мире.

Предположим теперь, что я еще не раскрыл существенных физических свойств воды (в действительном мире), т. е. я еще не знаю, что вода — это H_2O . Я могу располагать довольно надежными способами *распознавания* воды (конечно, я могу иногда и ошибиться, и ошибка выявится на более поздней стадии научного развития), но я могу не знать микроструктуры воды. Если признать, что жидкость с внешними свойствами «воды», но с другой микроструктурой *на самом деле не вода*, то мои методы распознавания воды (мое, так сказать, «операциональное определение» воды) нельзя считать аналитическим определением *того, что значит быть водой*. Скорее, операциональное определение, подобно остенсивному определению, представляет собой просто способ указания на образец, когда в *действительном мире* указывается такое вещество, что для x быть водой в *любом* мире, значит находиться в отношении «та же самая жидкость» к *нормальным* представителям класса *местных* (*local*) объектов, удовлетворяющих данному операциональному определению. «Вода» на Двойнике Земли не является водой, даже если она удовлетворяет операциональному определению, потому что она не находится в отношении *та же самая*_ж к тому местному веществу, которое удовлетворяет операциональному определению,

равно как не является водой и местное вещество, удовлетворяющее операциональному определению, но имеющее иную по сравнению с остальным местным веществом, удовлетворяющим операциональному определению, микроструктуру, потому что оно не находится в отношении «та же самая жидкость» к *нормальным* образцам местной «воды».

Предположим теперь, что я открыл микроструктуру воды, установив, что вода — это H_2O . Теперь я могу сказать, что то вещество на Двойнике Земли, которое я раньше *ошибочно принимал* за воду, на самом деле водой не является. Сходным образом, если вы описываете не другую планету в действительной вселенной, а другую возможную вселенную, где существует вещество с химической формулой XYZ , которое проходит «операциональный тест» для воды, то вы должны будете признать, что это вещество не вода, а XYZ . При этом вы описываете не возможный мир, где «вода — это XYZ »; вы описываете возможный мир, в котором озера содержат XYZ , люди пьют XYZ (а не воду) и т. п. Фактически, как только мы раскрыли природу воды, ни один мир, где вода не имеет такой природы не может рассматриваться как возможный. Как только мы открыли, что вода (в действительном мире) есть H_2O , ни один мир, где вода не есть H_2O , не может рассматриваться как возможный. В частности, если «логически возможное» утверждение — это утверждение, истинное в некотором «логически возможном мире», то утверждение «вода не есть H_2O » не является логически возможным.

С другой стороны, вполне можно представить себе, что опыт убедит нас (и придаст рациональный характер нашей вере в то), что вода не есть H_2O . В этом случае мыслимо, что вода не есть H_2O . Мыслимо, но не логически возможно! Быть мыслимым не значит быть логически возможным.

Крипке называет утверждения, которые рационально не могут быть пересмотрены (допустим, что такие существуют), *эпистемологически необходимыми*. Утверждения, истинные во всех возможных мирах, он определяет как просто необходимые (иногда — как «метафизически необходимые»). Используя эту терминологию можно переформулировать сказанное выше следующим образом: утверждение может быть (метафизически) необходимым и эпистемически случайным. Человеческая интуиция не имеет привилегированного доступа к метафизической необходимости.

Со времен Канта философы разделились на тех, кто считал все необходимые истины аналитическими, и тех, кто полагал, что некоторые необходимые истины являются синтетическими *a priori*. Но

ни один из этих философов не считал, что (метафизически) необходимая истина может не быть априорной: кантовская традиция вместе с эмпиристской традицией повинны в отождествлении метафизической и эпистемической необходимости. В этом смысле доводы Крипке против общепринятой доктрины выходят далеко за рамки обычного противопоставления эмпирицизма и кантианства.

Однако, в настоящей статье нас интересует теория значения, а не теория необходимой истины. Идеи, близкие идеям Крипке, были сформулированы с использованием понятия *индексальности* (indexicality)⁹. Такие слова, как «сейчас», «это», «здесь» издавна признаны *индексальными* или *контекстно-зависимыми* (token-reflexive), т. е. имеющими экстенционал, меняющийся от контекста к контексту или от одного употребления знака (token) к другому. Применительно к этим словам никто никогда не формулировал традиционного тезиса о том, что «интенционал определяет экстенционал». Возьмем наш пример с Двойником Земли: если у меня есть *Doppelgänger* на Двойнике Земли, то когда я думаю: «У меня болит голова», он тоже думает: «У меня болит голова». Однако экстенционалом слова «я», когда его употребляет в своей вербализованной мысли мой *Doppelgänger*, является он сам (а, точнее, множество, где он является единственным элементом), тогда как экстенционалом слова «я» в *моей* вербализованной мысли являюсь я (а, точнее, множество, где я — единственный элемент). Стало быть, одно и то же слово «я» имеет в двух разных идиолектах два разных экстенционала, но отсюда не следует, что понятие, которое у меня сложилось о себе, чем-либо отличается от того понятия, которое имеет мой *Doppelgänger*.

На наш взгляд, индексальность распространяется не только на явно индексальные слова и морфемы (например, временные показатели у глаголов). Нашу теорию можно подытожить следующим образом: слова типа «вода» имеют неявный (unnoticed) индексальный компонент: «вода» — это вещество, которое находится в определенном отношении подобия к воде *в данном месте*. Вода в другое время и в другом месте или даже в другом возможном мире, *чтобы быть водой*, должна находиться в отношении «та же самая жидкость» к *нашей* «воде». Таким образом, теория, согласно которой (1) слова имеют «интенционалы», представляющие собой что-то вроде понятий, которые говорящий связывает с этими словами, и (2) ин-

⁹ Эти идеи были сформулированы в моих лекциях, прочитанных в 1968 г. в Вашингтонском Университете и Университете шт. Миннесота.

тенсионал слова определяет его экстенционал, не может быть истинной теорией для таких терминов естественных видов как «вода» — по той же причине, по которой эта теория не может быть истинной для явно индексальных слов типа «я».

Однако теория, утверждающая индексальность терминов естественных видов типа «вода», оставляет открытым вопрос о том, считать ли, что «вода» в диалекте Двойника Земли имеет то же значение, что и «вода» в диалекте Земли, но другой экстенционал (именно так мы обычно и говорим о слове «я» в разных идиолектах). Тогда придется отказаться от тезиса о том, что «значение (интенционал) определяет экстенционал». Или же следует считать, — как мы это и делаем, — что для терминов естественных видов различие в экстенционалах *ipso facto* является различием в значении, и тем самым отвергнуть тезис о том, что значения представляют собой понятия или вообще *какие-то* ментальные сущности.

Впрочем, понятно, что и концепция Крипке, согласно которой термины естественных видов являются жесткими десигнаторами, и наша концепция об их индексальной природе — это лишь два разных способа выражения одной идеи. Мы полностью присоединяемся к словам Крипке:

«Предположим, что мы фиксируем референцию имени при помощи дескрипции. Поступая так, мы не делаем имя синонимом дескрипции, а употребляем имя как *жесткий десигнатор* для референции к называемому им объекту — даже при описании контрфактических ситуаций, где называемый этим именем объект данной дескрипции не соответствует. Полагаю, что для тех случаев, где референция фиксируется дескрипцией, это действительно справедливо. Но, в противоположность большинству современных теоретиков, я считаю, что референция имен редко или практически никогда не фиксируется с помощью дескрипций. Я под этим разумею не то, о чем у Сёрля сказано: "Референцию фиксирует не единичная дескрипция, а, скорее, пучок, семейство признаков". Я имею в виду, что признаки в этом случае совсем не используются»¹⁰.

¹⁰ Kripke S. "Identity and necessity" // Munitz M. (ed.) *Identity and Individuation*, N. Y., 1972, p. 135—164. В русск. переводе: Крипке С. Тожество и необходимость // Новое в зарубежной лингвистике. Вып. XIII. «Логика и лингвистика». М., 1982, с. 366.

БУДЕМ РЕАЛИСТАМИ!

Теперь я хотел бы сопоставить свою точку зрения с достаточно распространенным, по крайней мере среди ученых, воззрением (сложившимся, по всей видимости, совершенно стихийно). В качестве примера термина естественных видов возьмем слово *золото*. Для нас не существенны различия между словом «золото» и родственными ему словами в греческом, латинском и других языках, и мы ограничимся рассмотрением золота только в его твердом состоянии. Приняв все эти оговорки, мы утверждаем следующее: «золото» не изменило своего *экстенционала* (или не изменило его значительным образом) за последние две тысячи лет. За это время методы *идентификации* золота невероятно усложнились, но экстенционал слова *χρυσός* в греческом диалекте Архимеда совпадает с экстенционалом слова *золото* в моем английском диалекте.

Допустим (а это вполне могло быть), что как существовали кусочки металла, относительно которых до Архимеда нельзя было установить, что они не являются золотом, так существовали кусочки металла, относительно которых и во времена Архимеда нельзя было установить, что они не являются золотом, но сегодня с помощью современной техники мы легко можем отличить их от золота. Пусть *X* будет таким кусочком металла. Очевидно, что *X* не входит в экстенционал современного слова «золото»; моя позиция такова, что он не входил и в экстенционал древнегреческого слова *χρυσός*, хотя древний грек мог ошибочно принимать *X* за золото (или, скорее, за *χρυσός*).

Согласно альтернативной точке зрения слово «золото» *обозначает* все то, что удовлетворяет *имеющемуся на тот момент* «операциональному определению» *золота*. Сотню лет назад слово «золото» обозначало все то, что удовлетворяло принятому сто лет назад «операциональному определению» *золота*. В настоящий момент слово «золото» обозначает все то, что удовлетворяет операциональному определению золота, которое мы используем сейчас, в 1973 г., а слово *χρυσός* обозначало все то, что удовлетворяло операциональному определению *χρυσός*, применявшемуся *в то время*.

Принятие этой точки зрения обычно мотивируется определенным скептицизмом в понимании *истины*. Согласно моей точке зрения, когда Архимед утверждал, что некий кусочек металла есть золото (*χρυσός*), он утверждал не только то, что этот кусочек имеет внешние признаки золота (в исключительных случаях объект может принадлежать к естественному виду, *не имея* внешних признаков этого естественного вида); он утверждал, что этот кусочек имеет та-

кую же *внутреннюю* (hidden) *структуру* (иными словами, такую же «сущность»), какую имеет любой обычный кусочек местного золота. Архимед признал бы наш гипотетический кусочек металла X золотом, но он был бы *неправ*. Но *кому решать*, что он был *неправ*?

Очевидный ответ: *нам* (имеющим в своем распоряжении сегодня более совершенную теорию). Для большинства людей или вопрос (кому решать?) правомерен, а наш ответ — нет, или наш ответ правомерен, а сам вопрос — нет. Почему это так?

Думаю, причина в том, что в своих интуитивных представлениях люди склонны быть или непреклонными антиреалистами или непреклонными реалистами. Для непреклонного антиреалиста утверждение, что принадлежность того или иного объекта к экстенционалу архимедовского термина *χρυσός* следует устанавливать с помощью *нашей* теории, имеет немного смысла. Антиреалист не считает нашу теорию и теорию Архимеда двумя приблизительно правильными описаниями некоторой фиксированной области независимых от теории сущностей и скептически относится к идее «конвергенции» в науке, а потому для него наша теория не выступает *лучшим* описанием *тех же самых* сущностей, которые описывал Архимед. Но, если наша теория — это *лишь* наша теория, то попытка с *ее* помощью решать, входит ли X в экстенционал слова *χρυσός* или нет, столь же *неоправданна*, как и попытка решать этот вопрос с помощью теории неандертальцев. Единственная теория, для использования которой есть оправдание, — это теория, которой придерживается сам носитель языка.

Трудность состоит в том, что для непреклонного антиреалиста *истина* имеет смысл только как внутритеоретическое понятие. Антиреалист может использовать истину как внутритеоретическое понятие, трактуя ее как «избыточную теорию» (redundancy theory); но ему недоступны *внетеоретические* понятия истины и референции. Однако *экстенционал связан с понятием истины*. Экстенционалом термина является именно то, относительно *чего* термин *истинен*. Вместо того, чтобы пытаться сохранить понятие экстенционала при помощи нелепого операционализма, антиреалист должен отказаться от понятия экстенционала так же, как он отказался от понятия истины (в любом внетеоретическом смысле). Например, подобно Дьюи, он может взамен истины воспользоваться понятием «правомерной утверждаемости» (warranted assertibility) (определяемой относительно научного метода как такового, если он считает, что существует *неизменный* научный метод, или относительно наилучших

методов, доступных в данное время, если он согласен с Дьюи, что сам научный метод развивается). Тогда он может говорить, что во времена Архимеда было правомерно утверждать, что «*X* есть золото (*χρυσός*)», а сегодня это неправомерно (по сути, это *минимальное* требование, относительно которого могут согласиться реалист и антиреалист). Однако утверждение, что *X* входил в экстенционал слова *χρυσός*, будет отброшено как бессмысленное вместе с утверждением, что высказывание «*X* есть золото (*χρυσός*)» было *истинно*.

Хорошо известно, что узкий операционализм не может объяснить реальное употребление научных терминов и терминов обыденного языка. Более слабые версии операционализма типа карнаповской версии теории Рамсея, если и не объясняют, то согласуются с реальным употреблением научных терминов (главным образом потому, что более слабые версии согласуются с любым возможным употреблением!), — достигается это ценой превращения наследуемости научных результатов в *чудо*. Нет никаких сомнений в том, что ученые используют термины так, как если бы связанные с этими терминами критерии были не *необходимыми* и *достаточными* условиями, а скорее *приблизительно* верными характеристиками некоторого мира независимых от теории сущностей, и рассуждают так, как если бы более поздние теории зрелой науки были, в целом, *лучшими* описаниями *тех же самых* сущностей, на которые ссылаются более ранние теории. На мой взгляд, только та гипотеза, которая признает, что это *так и есть*, может объяснить наследование научных результатов, ограничение приемлемых научных теорий логикой первого порядка и многие другие характеристики научного метода¹¹. Но в мои задачи здесь не входит обоснование этого. Мой тезис таков: если нам приходится использовать истину и экстенционал как внетеоретические понятия (т. е. рассматривать их как понятия, применяемые к высказываниям, сформулированным на языках, отличных от языка нашей теории), то нам следует принять реалистическую перспективу, к которой эти понятия и относятся. Сомневаться относительно того, можем ли мы утверждать, что *X* не входит в экстенционал слова «золото», когда его употребляет Джон, — это все равно что сомневаться относительно того, имеет ли смысл считать высказывание Джона, что «*X* есть золото», *истинным или ложным* (а не просто «правомерно утверждаемым» Джоном, но «не-

¹¹ Великолепное обсуждение именно этих проблем см. в работе: Boyd R., *Realism and Scientific Epistemology* (неопубликована, ксерокопия рассылается автором, отделение философии Корнельского университета).

правомерно утверждаемым» нами). Попытка согласовать понятие истины, реалистическое по своей сути, с антиреалистическими предрассудками, прибегнув при этом к неадекватной теории значения, не может иметь успеха.

Другой причиной, побуждающей принять крайний операционализм, служит неприязнь к неverified гипотезам. Как может показаться на первый взгляд, мы утверждаем, что высказывание « X есть золото ($\chi\rho\upsilon\sigma\omicron\varsigma$)» было ложным во времена Архимеда, хотя Архимед *в принципе* не мог знать, что оно ложно. Но это не совсем так. На деле, мы можем описать множество ситуаций (воспользовавшись теорией, утверждающей, что X не есть золото), в которых X проявлял бы себя совсем иначе, чем остальное вещество, определяемое Архимедом как золото. Возможно, X разлагался бы на два разных металла при плавлении, имел бы другую проводимость или превращался бы в газ при другой температуре и т. п. Если бы мы выполнили эти эксперименты в присутствии Архимеда, то он, не зная указанной теории, смог бы установить ту эмпирическую регулярность, что « X ведет себя в некоторых отношениях иначе, чем остальное вещество, которое я определяю как $\chi\rho\upsilon\sigma\omicron\varsigma$ ». В конце концов он сделал бы вывод, что «Возможно, X не есть золото».

Суть в том, что даже если что-то удовлетворяет используемым в данное время критериям, позволяющим идентифицировать золото (т. е. установить, является ли нечто золотом), в одной или нескольких ситуациях оно проявляет себя иначе, чем остальное вещество, удовлетворяющее этим критериям. Возможно, это не *доказывает*, что оно не является золотом, но это оправдывает гипотезу, утверждающую, что оно может не быть золотом, даже если этому нет теоретического объяснения. Если бы мы сообщили Архимеду, что золото, в отличие от X , имеет такую-то молекулярную структуру, а металл X ведет себя иначе, потому что имеет другую молекулярную структуру, можем ли мы сомневаться в том, что Архимед согласился бы с нами, и что X — это не золото? В любом случае, на мой взгляд, нелепо испытывать беспокойство по поводу того, что нечто может быть *истинным* (в данный момент времени), но в то же время не может быть *верифицировано*. Согласно любой разумной точке зрения, безусловно, существуют истинные вещи, которые *никогда* нельзя будет верифицировать. Например, предположим, что существует бесконечное множество двойных звезд. *Должны ли мы быть способными верифицировать этот факт, хотя бы в принципе?*

До сих пор мы имели дело с *метафизическими* основаниями, выдвигаемыми против нашей точки зрения. Но с нами могут не со-

гласиться и в отношении эмпирических фактов о намерениях говорящих. Например, кто-то может считать, что Архимед (в описанном выше *Gedankenexperiment*¹²) сказал бы: «неважно, что X проявляет себя иначе, чем остальные кусочки золота; X является кусочком золота, потому что обладает такими-то и такими-то свойствами, а их достаточно, чтобы быть золотом». Хотя мы действительно не можем быть уверенными в том, что слова, обозначающие естественные виды в древнегреческом языке, выражали те же свойства, что и соответствующие слова в современном английском, у нас не может быть никаких сомнений относительно свойств, выражаемых последними. Если мы отставим в сторону философские предубеждения, то, думаю, нам станет совершенно ясно, что никакое операциональное определение не обеспечивает необходимого и достаточного условия для применения любого такого слова. Мы можем дать «операциональное определение», указать пучок признаков и т. п., но в наше намерение никогда не входит «сделать имя синонимом дескрипции». Скорее «мы употребляем имя как жесткий десигнатор» для обозначения различных вещей, имеющих ту природу, которую обычно имеют вещи, удовлетворяющие данной дескрипции.

ДРУГИЕ СМЫСЛЫ

До сих пор мы анализировали основной смысл терминов естественных видов (или, скорее, их основной *экстенционал*). Однако термины естественных видов, как правило, обладают некоторым множеством смыслов. (Зифф даже предположил, что они обладают *континуумом* смыслов).

Отчасти этот факт можно объяснить на основе нашей теории. Например, для того, чтобы быть водой, нужно находиться в отношении «та же самая жидкость» к определенным вещам. Но что это за отношение «та же самая жидкость»?

X находится в отношении «та же самая жидкость» к y только в том случае, если (1) x и y являются жидкостями, и (2) x и y совпадают в своих важных физических свойствах. Термин «жидкость» сам является термином естественных видов, но я не буду его здесь анализировать. Термин «свойство» является термином с широким спектром значений, и мы проанализировали его в предыдущих своих работах. В настоящей статье я хотел бы сосредоточить свое внимание на понятии «важность». Важность зависит от интереса. Как прави-

¹² Мысленный эксперимент (нем. яз.) — прим. перев.

ло, «важными» свойствами жидкости, твердого вещества и т. п. являются свойства, важные в *структурном* отношении, т. е. свойства, фиксирующие, что жидкость, твердое вещество и т. п. в конечном счете состоят из элементарных частиц, водорода и кислорода, земли, воздуха, огня, воды или т. п., и определяющие, каким образом они из них составлены и могут поэтому проявлять соответствующие внешние признаки. С этой точки зрения характерной особенностью какого-либо объема воды обычно является то, что он состоит из молекул H_2O . Однако иногда может оказаться важным наличие примесей; так, в одном контексте «вода» может означать *химически чистую воду*, а в другом контексте она может означать жидкость в озере Мичиган. Кто-то может иногда говорить о жидкости XYZ как о воде, если он ее *использует* в качестве воды. Кроме того, обычно важно, что вода находится в жидком состоянии; но иногда это неважно и можно называть водой одну-единственную молекулу H_2O или водяные пары.

Даже смыслы настолько экстравагантные, что их следует рассматривать как «девиантные», могут иметь определенное отношение к основному смыслу. Например, я могу спросить: «Видели ли Вы лимон?», имея в виду *пластиковый* лимон. А вот не столь девиантный случай: Мы обнаружили на Марсе «тигров», то есть, они выглядят как тигры, но в основе их химического состава лежит не углерод, а кремний. (Замечательный пример параллельной эволюции!) Являются ли марсианские «тигры» тиграми? Это зависит от контекста.

В случае этой теории, как и в случае любой другой теории, идущей вразрез с тем, что люди привыкли думать о каком-то предмете, неизбежно возникают неправильные истолкования. Одно из таких неправильных истолкований уже появилось. Так, один критик утверждает, что *основным* является тот смысл слова, скажем, слова «лимон», в каком все, что обладает внешними признаками лимона (или достаточным количеством этих признаков), является лимоном. По мнению того же критика, чтобы быть лимоном, предмету нужно обладать скрытой структурой — генетическим кодом лимона — лишь в том случае, если слово «лимон» используется как *научный* термин. Представляется, что обе эти точки зрения основываются на неправильном истолковании или, возможно, на двух взаимодополняющих неправильных истолкованиях.

Тот смысл, в каком буквально *все*, что имеет внешние признаки лимона, с необходимостью является лимоном, представляет собой вовсе не основной, а крайне девиантный смысл. В этом смысле любой предмет был бы лимоном, имей он внешний вид и вкус лимона,

даже если, к примеру, его химический состав был бы основан на кремнии, или с помощью электронного микроскопа было бы установлено, что он является *машиной*. (Включив в число внешних признаков тот факт, что он рос «как лимон», мы не исключим кремниевый лимон, если на Марсе существуют «лимонные» деревья. Этим мы не исключим и машину-лимон, поскольку дерево тоже может быть машиной!)

Вместе с тем смысл, в каком предмету, чтобы быть лимоном, нужно иметь генетический код лимона, *не тождественен* техническому смыслу (если таковой имеется, в чем я сомневаюсь). В техническом смысле, как я его понимаю, слово «лимон» было бы *синонимом* дескрипции, *специфицирующей* генетический код. Но когда мы говорим (если сменить пример), что веществу нужно быть H_2O , чтобы быть *водой*, мы не имеем в виду, что *носитель языка* должен это *знать*. Если быть H_2O — это (метафизически необходимое) условие истинности для того, чтобы быть *водой*, то, только подменяя *метафизическую* необходимость *эпистемологической*, мы можем заключить, что слово «вода» должно быть синонимично H_2O , и в этом случае оно определено выступает научным термином. Сходным образом, даже если основным для слова «лимон» является тот смысл, в каком предмету, чтобы быть лимоном, нужно (как я полагаю) иметь генетический код лимона, из этого не следует, что слово «лимон» является синонимом дескрипции, которая явным или неявным образом специфицирует генетический код.

Мы еще более убедимся в ошибочности того мнения, что есть важный (возможно, основной) смысл слова «лимон», в каком предмету, чтобы быть лимоном, по крайней мере, *достаточно* обладать внешними признаками лимона, если в набор внешних признаков включим способность к перекрестному опылению с лимонами. Признак перекрестного опыления с лимонами предполагает понятие лимона. Поэтому, даже если *таким* образом можно сформулировать достаточное условие, но считать это несовместимым с предложенным здесь описанием — значит уходить от вопроса. Более того, описание в терминах предполагаемых *лимоном* «внешних признаков» (наподобие перекрестного опыления с *лимонами*) не задает условий истинности, которые позволили бы нам определить, какие объекты в возможных мирах (или какие объекты миллионы лет назад, или какие объекты на расстоянии в миллион световых лет отсюда) являются лимонами. (К тому же, я не считаю это описание, фактически уводящее нас от существа дела, *корректным*, даже в качестве достаточного условия. Полагаю, что можно придумать слу-

чай, когда то, что не является лимоном, перекрестно опыляется с лимоном и выглядит как лимон и т. п.)

Можно было бы попытаться исключить случай машины-лимона (или лимона-машины?), которая «растет» на машине-дереве (или дереве-машине?), сказав, что подобный «рост» на самом деле не является *ростом*. Это верно; но это верно потому, что глагол «расти» — это *термин* естественных видов, и к *нему* применяется именно тот вид объяснения, который мы здесь предложили.

Следует избегать еще одного неправильного истолкования, которое состоит в следующем: считается, что наше объяснение предполагает, что элементы, входящие в экстенционал термина естественного вида, с необходимостью *обладают* общей внутренней структурой. Ведь могло бы оказаться, что некоторые объемы жидкости, которую мы называем «водой», помимо внешних признаков *не* имеют *никаких* других важных общих физических характеристик. В этом случае необходимым и достаточным условием того, чтобы быть «водой», было бы обладание достаточным количеством внешних признаков.

Кстати, последнее утверждение не предполагает, что вода могла бы не иметь внутренней структуры (или что вода могла быть не H_2O , а чем-то иным). Когда мы говорим, что могло бы *оказаться*, что вода не имеет внутренней структуры, мы имеем в виду, что жидкость без какой-либо внутренней структуры (т. е. множество объемов других жидкостей, не имеющих ничего общего, *кроме* внешних признаков), могла бы выглядеть как вода, иметь вкус воды, заполнять озера и т. п., которые в действительности заполнены водой. Короче говоря, мы могли бы находиться в том же самом эпистемологическом состоянии по отношению к жидкости без внутренней структуры, в каком мы когда-то находились по отношению к воде. Сопоставьте это с примером Крипке об «аналое, сделанном из льда»¹³.

В реальности присутствует почти весь спектр случаев. Например, некоторые болезни не имеют общей внутренней природы (единственное, что их роднит, — это единый пучок симптомов), тогда как другие болезни имеют общую этиологию и природу (например, разные виды туберкулеза). В ряде случаев нам еще не удалось этого установить, а в отношении рассеянного склероза до сих пор не стихают споры.

¹³ См. сноску 10 — *Прим. перев.*

Интересный случай представляет *нефрит*. Хотя китайцы не признают этого различия, но слово «нефрит» обозначает два разных минерала: жадеит и нефрит. Они явным образом различаются по своему химическому составу. Жадеит представляет собой соединение натрия и алюминия. Нефрит же состоит из кальция, магния и железа. Однако эти совершенно разные микроструктуры порождают внешнюю текстуру с одними и теми же уникальными свойствами.

Вернемся на секунду к примеру с Двойником Земли; если бы и H_2O , и XYZ были в достаточном количестве на Земле, то мы бы имели случай, аналогичный жадеиту и нефриту: было бы корректно утверждать, что существует *два вида «воды»*. Мы бы не говорили тогда, что «вещество на Двойнике Земли на деле оказалось не водой», а сказали бы, что «оно оказалось тем *видом воды*, который имеет формулу XYZ ».

Итак: если существует внутренняя структура, то, как правило, она и определяет принадлежность к естественному виду не только в действительном мире, но и во всех возможных мирах. Иными словами, она определяет то, что мы можем и что не можем контрфактически предполагать относительно естественного вида («могла бы вся вода испариться?» — «да»; «могла бы вода быть XYZ » — «нет»). Однако местная вода или любое другое вещество могла бы иметь две или более внутренние структуры, или так много «внутренних структур», что это понятие утратило бы значение, а решающими стали бы тогда внешние признаки.

ДРУГИЕ СЛОВА

До сих пор мы использовали в качестве примеров только термины естественных видов, но наши положения применимы и к многим другим категориям слов. Они применимы к подавляющему большинству существительных, равно как и к другим частям речи.

Рассмотрим, например, имена созданных человеком предметов, т. е. слова типа «карандаш», «стул» «бутылка» и т. п. Согласно традиционной точке зрения эти слова точно определяются конъюнкциями, или пучками, признаков. Предмет, обладающий всеми признаками, входящими в конъюнкцию (или достаточным количеством свойств, образующих пучок), с необходимостью является *карандашом, стулом, бутылкой* и т. п. Кроме того, в случае пучка признаков *необходимыми* обычно считаются некоторые признаки в пучке (в случае конъюнкции признаков, *все* признаки, входящие в конъюнкцию, являются *необходимыми*). Свойство *быть созданным человеком* име-

ет предположительно необходимый характер, а свойство «принадлежать к виду, имеющему некоторое стандартное назначение» считается необходимым, как например, в предложениях «карандаши созданы человеком» и «карандаши стандартно предназначены для письма». Наконец, этот вид необходимости трактуется как необходимость в эпистемическом смысле, а фактически — и в аналитическом смысле.

Давайте еще раз обратимся к научной фантастике. На сей раз мы воспользуемся примером, придуманным Роджером Альбриттоном. Представьте себе, в один прекрасный день мы делаем открытие, что *карандаши являются организмами*. Мы разрезаем их, изучаем под электронным микроскопом и обнаруживаем почти невидимое сплетение нервов и других внутренних органов. Мы ведем за ними слежку и устанавливаем, что они размножаются, а их потомство вырастает и становится взрослыми карандашами. Мы обнаруживаем, что эти организмы не имитируют другие (созданные человеком) карандаши; нет и никогда не было других карандашей, кроме этих организмов. Конечно, странно, что на многих этих организмах есть надписи, например, BONDED GRANTS DELUXE made in U.S.A. No. 2, — возможно, как разумные существа, они используют это в качестве камуфляжа. (Нам также придется объяснить, потому никто никогда даже не пытался производить карандаши и т. д., но ясно, что в некотором смысле это можно считать возможным миром.)

Если это возможно, а я согласен с Альбриттоном, что это возможно, то значит эпистемологически возможно, что *карандаши могли бы оказаться организмами*. Отсюда следует, что утверждение *карандаши создаются человеком* не является эпистемологически необходимым в строгом смысле и, *a fortiori*, не является и аналитически истинным.

Однако будем осторожными. Доказали ли мы этим, что есть возможный мир, в котором карандаши являются организмами? Думаю, что нет. Мы доказали только, что есть возможный мир, в котором определенные организмы *являются эпистемическими двойниками* карандашей (в терминологии Крипке). Возвратимся к примеру с Двойником Земли и представим на сей раз, что карандаши на Земле — это, как мы и считаем, создаваемые человеком приспособления для письма, тогда как «карандаши» на Двойнике Земли — это организмы а-ля Альбриттон. Далее представим себе, что жители Двойника Земли совершенно не подозревают об этом и имеют о «карандашах» такие же представления, что и мы — о карандашах. Обнаружив это, мы не стали бы говорить: «некоторые карандаши —

это организмы». Скорее всего мы бы сказали: «вещи, которые на Двойнике Земли, принимают за карандаши, на самом деле карандашами не являются. В действительности они представляют собой разновидность организмов».

Предположим теперь, что и на Земле, и на Двойнике Земли ситуация такова, как ее описал Альбриттон. Тогда мы бы сказали: «карандаши — это организмы». Таким образом, являются ли «карандаши» на Двойнике Земли (или в другой возможной вселенной) *карандашами* или нет, — зависит от того, являются ли *местные* карандаши организмами или нет. Если местные карандаши именно таковы, какими мы их считаем, то возможный мир, в котором есть карандаши-организмы, — это *не тот* возможный мир, где *карандаши являются организмами*; в этом случае (а он, конечно же, и имеет место), *не существует* возможных миров, в которых карандаши являются организмами. Утверждение, что карандаши создаются людьми, является необходимым в том смысле, что оно истинно во всех возможных мирах, т. е. оно является метафизически необходимым. Однако отсюда не следует, что оно эпистемически необходимо.

Отсюда следует, что «карандаши» — это не синоним какой-либо дескрипции — и даже не приблизительный синоним *приблизительной* дескрипции. Когда мы употребляем слово «карандаш», мы имеем в виду то, что обладает такой же *природой*, как и обычные образцы местных карандашей в действительном мире. Слово «карандаш» так же *индексально*, как и слово «вода» или «золото».

В известной мере пример о карандашах, оказавшихся организмами, дополняет пример, предложенный нами несколько лет назад¹⁴, — о кошках, оказавшихся роботами (и дистанционно управляемых с Марса). В своей подготавливаемой к печати работе Кац утверждает, что мы неправильно охарактеризовали случай с кошками; его следовало бы описать так: *оказалось, что в этом мире нет кошек*. Кац допускает, что мы могли бы сказать: «Кошки оказались не животными, а роботами»; но он утверждает, что это — семантически девиантное предложение, которое переводится так: «Вещи, которые я обозначаю словом “кошки”, оказались не животными, а роботами». Однако теория Каца — это плохая лингвистика. Прежде всего, его объяснение, почему мы можем *сказать* «Кошки являются роботами», — это просто объяснение того, как вообще мы можем что-либо сказать. Важнее то, что по теории Каца предложение «Кошки явля-

¹⁴ Putnam H. It ain't necessarily so // Putnam H. Mathematics, Matter and Method. Philosophical Papers, Cambridge..., 1975, Vol. I, p. 237—249.

ются роботами» *девиантно*, тогда как предложение «В этом мире не существует кошек» — недевиантно, а, по сути, стандартно для описываемого случая. Я не отрицаю, что в *каком-то* случае предложение «В мире не существует (и никогда не существовало) никаких кошек» было бы стандартным: Мы могли бы (в эпистемическом плане) установить, что страдали коллективной галлюцинацией. («Кошки выступают здесь чем-то вроде чёртиков.») Но в описываемом мною случае предложение «Кошки оказались роботами, дистанционно управляемыми с Марса» определенно является недевиантным, а предложение «В мире не существует кошек» является в высшей степени девиантным.

Кстати, объяснение Каца — это не только плохая лингвистика; оно не годится и в качестве рациональной реконструкции. Безусловно, причина, почему мы *не* используем слово «кошка» как синоним некоторой дескрипции заключается в том, что мы знаем достаточно о кошках и знаем, что они обладают внутренней природой, а хорошая научная методология состоит в том, чтобы использовать имя как жесткий десигнатор тех вещей, которые обладают этой внутренней природой, а вовсе не тех, которые, как оказывается, удовлетворяют некоторой дескрипции. Конечно, если бы мы *знали* внутреннюю структуру, мы могли бы сформулировать дескрипцию в терминах *этой внутренней структуры*; но мы пока этого не делаем. В этом смысле употребление терминов естественных видов говорит нечто важное о нашем отношении к миру: мы знаем, что существуют виды вещей, обладающих общей внутренней структурой, но у нас нет еще знаний для описания всех этих внутренних структур.

Однако, позиция Каца в случае с «карандашами» более правдоподобна, чем в случае с «кошками». Мы полагаем, что мы *знаем* необходимое и достаточное условие того, чтобы быть *карандашом*, хотя оно и очень неопределенно. Поэтому слово «карандаш» может быть синонимом некоторой приблизительной дескрипции. В случае, когда «карандаши оказались организмами», мы *могли бы* сказать, что «карандаши оказались организмами» или же что «в мире не существует карандашей», т. е. мы могли бы употребить слово «карандаш» или как термин естественного вида или как «однокритериальное» слово ¹⁵.

¹⁵ Идея «однокритериального» слова и теория аналитической истины, основанная на этом понятии, представлена в статье: Putnam H. The Analytic and the Synthetic // Putnam H. Mind, Language and Reality. Philosophical Papers, Cambridge..., 1975, Vol. 2, p. 33—69.

С другой стороны, мы могли бы поставить под сомнение, что в естественном языке вообще *есть по-настоящему* однокритериальные слова, если не считать контекстов, где это специально оговорено. Разве не могло бы так оказаться, что педиатры — это не врачи, а шпионы с Марса? Ответив «да», вы отбрасываете синонимичность выражений «педиатр» и «врач, специализирующийся по уходу за детьми». Представляется, что слова, введенные как «однокритериальные», имеют сильную тенденцию становится в некотором смысле терминами «естественных видов», со всей характерной для них жесткостью и индексальностью. В случае имен создаваемых человеком предметов эта тенденция, видимо, затрагивает основной смысл этих слов.

(Есть шутка о пациенте, которого вот-вот должны выписать из психиатрической больницы. Врачи в течение какого-то времени задают ему вопросы, и он совершенно разумно на них отвечает. Они решают его выписать, но в конце беседы один из врачей случайно спрашивает: «Кем вы хотите быть, когда выйдете из больницы?» — «Чайником». Шутка не была бы остроумной, если бы о человеке в переносном смысле нельзя было сказать, что он — чайник).

Однако есть слова, сохраняющие практически в чистом виде однокритериальный характер. Это слова, значение которых выражается таким преобразованием: *охотник = тот, кто охотится*.

Предложенное нами объяснение применимо не только к большинству существительных, но и к другим частям речи. Глаголы типа «расти», прилагательные типа «красный» и т. д., все имеют индексальный характер. С другой стороны, некоторые синкатегорематические слова, видимо, в большей мере имеют однокритериальный характер. Слово «весь», например, можно объяснить следующим образом: предложение «Армия окружила город» может быть истинным, даже если дивизия А не принимала в этом участия. Предложение «Вся армия окружила город» означает, что каждое армейское подразделение (соответствующего вида, например, дивизия А) приняло участие в действии, выраженном глаголом¹⁶.

¹⁶ Этот пример взят из докторской диссертации Энтони Кроха (защищенной в Массачусетском технологическом институте на отделении лингвистики в 1974 г.)

ЗНАЧЕНИЕ

Определим теперь нашу позицию в отношении понятия значения. Мы установили, что экстенционал термина не фиксируется с помощью понятия, которое индивидуальный носитель языка имеет в своей голове, и это верно как в силу того, что экстенционал обычно определяется *социально* (существует разделение лингвистического труда в такой же мере, как и разделение «реального» труда), так и в силу того, что экстенционал отчасти определяется *индексально*. Экстенционал наших терминов зависит от реальной природы конкретных вещей, служащих образцами ¹⁷, и эта реальная природа обычно не известна в полной мере говорящему. Традиционная семантическая теория не учитывает всего лишь два фактора в определении экстенционала: фактор общества и фактор реального мира!

Вначале мы установили, что значение не может быть тождественно экстенционалу. Но оно не может быть тождественно и «интенционалу», если интенционал представляет собой что-то вроде *понятия* в голове индивидуального носителя языка. Что же тогда делать?

Есть два пути, которыми мы могли бы пойти. Первый путь — сохранить отождествление значения с понятием и отказаться от идеи, что значение определяет экстенционал. Если бы мы последовали этим путем, то мы могли бы сказать, что на Земле и на Двойнике Земли слово «вода» имеет одно и то же значение, но разные *экстенционалы*. (Не просто разные местные экстенционалы, но разные глобальные экстенционалы.) Жидкость XYZ на Двойнике Земли не входит в экстенционал произносимого мной слова «вода», но она входит в экстенционал слова «вода», когда его произносит мой *Doppelgänger*, но вовсе не потому, что Двойник Земли находится очень далеко от меня, ибо молекулы H_2O входят в экстенционал произносимого мной слова «вода», как бы далеко они ни находились от меня во времени и в пространстве. Кроме того, то, что я могу контрфактически предположить относительно воды, будет отличаться от того, что может относительно «воды» контрфактически предположить мой *Doppelgänger*.) Хотя это совершенно корректный подход к такому *абсолютно* индексальному слову, как «я», он представляется неправильным в отношении обсуждаемых нами слов. Рассмотрим, например, слова «вяз» и «бук». Если на Двойнике Зем-

¹⁷ Я не имею в виду понятия «образца», согласно которой любой образец *K* с необходимостью есть *K* (в реальности).

ли они «поменялись местами», то, конечно же, мы *не* сказали бы, что слово «вяз» имеет одно и то же значение на Земле и на Двойнике Земли, даже если имеющийся у моего *Doppelgänger*'а стереотип бука (или «вяза», как он его называет) тождествен моему стереотипу вяза. Скорее, мы сказали бы, что слово «вяз» в идиолекте моего *Doppelgänger*'а означает *бук*. По этой причине, видимо, предпочтительнее выбрать второй путь и отождествить «значение» с упорядоченной парой (или, возможно, с упорядоченной *n*-кой) сущностей, *одна из которых является экстенсионалом*. (Другие компоненты этого, так сказать, «вектора значения» будут определены ниже). Если выбрать этот путь, то станет тривиально истинным утверждение, что *значение определяет экстенсионал* (т. е. различие в экстенсионале является *ipso facto* различием в значении), но вместе с тем придется полностью отказаться от идеи, что при наличии различия в значении, которое мой *Doppelgänger* и я приписываем некоторому слову, *должно* быть некоторое различие и в наших понятиях (или в наших психологических состояниях). Следуя этим путем, мы можем сказать, что мой *Doppelgänger* и я *имеем в виду разное*, когда говорим «вяз», однако это ничего не говорит о наших психологических состояниях. Все это лишь означает, что когда это слово произносит он, оно имеет другой экстенсионал, чем когда его произношу я; но это различие в экстенсионалах не отражает какого-либо различия в нашем индивидуальном знании языка, взятом отдельно от всего остального.

Если это верно, а я думаю, что это так, то традиционная проблема значения распадается на две проблемы. Первая проблема — это объяснить, как *определяется экстенсионал*. Поскольку во многих случаях экстенсионал определяется не индивидуально, а социальным, вследствие разделения лингвистического труда, то я полагаю, что эта проблема является чисто социолингвистической проблемой. Для ее решения нужно детально сформулировать, каким именно образом осуществляется разделение лингвистического труда. Так называемая «каузальная теория референции», введенная Крипке для имен собственных и распространенная нами на термины естественных видов и физических величин¹⁸, относится к этой области исследований. Поскольку во многих контекстах мы приписываем произносимому нами имени ту референцию, которую мы приписываем этому же имени, когда оно произносится человеком, от которого мы

¹⁸ Putnam H. Explanation and Reference // Putnam H. Mind, Language and Reality. Philosophical Papers, Cambridge..., 1975, Vol. 2, p. 196—214.

узнали это имя (таким образом референция передается от говорящего к говорящему, начиная с тех, кто присутствовал «на церемонии именования», хотя никакой фиксированной *дескрипции* не передается), то это просто представляет собой конкретный случай социального сотрудничества при определении референции.

Другая проблема — это описать *индивидуальное знание языка* (*individual competence*). Экстенционал во многих случаях может определяться социально, но мы не приписываем стандартный экстенционал слову *W*, когда его произносит Джон, *не учитывая, как Джон употребляет слово W*. Джон должен иметь некоторые конкретные представления и навыки в отношении слова *W* для выполнения своей функции в разделении лингвистического труда. Поскольку мы отказались считать индивидуальное знание языка столь значительным, чтобы оно определяло экстенционал, то мы можем начать исследовать его с новой точки зрения.

В этой связи стоит отметить, что существительные типа «тигр» или «вода» весьма отличаются от имен собственных. Можно правильно использовать имя собственное «Сандерс», ничего не зная о его референте, за исключением того, что его зовут «Сандерс» — и даже это может не быть верным. («Давным-давно — кажется, в прошлую пятницу — Винни-Пух жил в лесу один-одинешенек, под именем Сандерс»). Однако нельзя правильно использовать слово тигр, кроме как *per accidens*, не зная довольно многих вещей о тиграх, или, по крайней мере, не будучи знакомым с определенным представлением о тиграх. В этом смысле понятия тесно *связаны* со значением.

Если исследование первой проблемы составляет тему для социолингвистики, то исследование второй проблемы образует тему для психолингвистики. К этой теме мы сейчас и обратимся.

СТЕРЕОТИПЫ И КОММУНИКАЦИЯ

Предположим, говорящий знает только, что слово «тигр» имеет в качестве экстенционала множество физических объектов, но не более того. Если он владеет обычными познаниями о других аспектах языка, то он мог бы использовать слово «тигр» в *некоторых* предложениях: например, «тигры имеют вес», «тигры занимают определенное пространство», «дайте мне тигра», «это тигр?» и т. п. Более того, *социально определяемый* экстенционал в этих предложениях был бы стандартным, т. е. был бы множеством тигров. Тем

не менее, мы не стали бы считать, что этот человек «знает значение» слова «тигр». Почему?

Прежде чем попытаться ответить на этот вопрос, давайте слегка его переформулируем. Мы будем говорить, что человек усвоил слово «тигр», если он способен употреблять его таким образом, что: (1) его употребление считается удовлетворительным (т. е. люди не говорят об этом человеке таких вещей, как: «он не знает, что такое тигр», «он не знает значения слова “тигр”» и т. п.); и (2) его общий способ обустроенности в мире и в лингвистическом сообществе таков, что социально определяемым экстенсионалом слова «тигр» в его идиолекте является множество тигров. Условие (1), по сути, означает, что мы не считаем носителя языка наподобие того, о котором шла речь в предыдущем абзаце, усвоившим слово «тигр» (или какое-либо другое слово). В некоторых случаях мы можем сказать, что он частично усвоил слово, но об этом немного позже. Условие (2) означает, что жители Двойника Земли, имеющие те же самые лингвистические привычки (*linguistic habits*), что и мы, считаются усвоившими слово «тигр» только в том случае, если экстенсионалом слова «тигр» в их идиолекте является множество тигров. Согласно основной идее в предыдущих разделах настоящей статьи из простой тождественности лингвистических привычек на Земле и на Двойнике Земли не следует, что экстенсионалом слова «тигр» в диалекте (или идиолектах жителей) Двойника Земли является множество тигров: здесь также имеет значение природа «тигров» на Двойнике Земли. (Если, к примеру, химический состав организмов на Двойнике Земли имеет кремниевую основу, то их «тигры» в действительности не являются тиграми, даже если они выглядят как тигры, а лингвистические привычки простого жителя Двойника Земли точно совпадают с лингвистическими привычками жителей Земли.) Таким образом, условие (2) означает, что в этом случае мы будем говорить, что жители Двойника Земли не усвоили нашего слова «тигр» (хотя они усвоили другое слово, которое точно так же пишется и произносится).

Введение выражения «усвоить слово» обусловлено тем, что вопрос «знает ли этот человек значение слова “тигр”?» неявно предполагает теорию, согласно которой усвоить слово значит знать то, что называют его «значением». Отождествив мы это значение с понятием, и мы вновь имеем теорию, утверждающую, что достаточным условием усвоения слова является соотнесение его с правильным понятием (или, вообще говоря, пребывание в правильном психологическом состоянии), т. е. мы вновь имеем ту самую теорию, кото-

рую все это время пытались опровергнуть. Поэтому далее мы будем «усваивать» слова, а не «знать их значения».

Теперь мы можем переформулировать вопрос, с которого начинается этот раздел. Употребление слова «тигр» описанным нами человеком не считается удовлетворительным, вместе с тем оно не вынуждает нас приписать нестандартный экстенционал слову «тигр» в идиолекте этого человека. Почему же оно не считается удовлетворительным?

Предположим, что наш гипотетический носитель языка указывает на снежный ком и спрашивает: «Это тигр?». Ясно, что нет смысла разговаривать с *ним* о тиграх. Осмысленная коммуникация предполагает, что люди хоть что-то знают о том, о чем они говорят. Разумеется, мы ежедневно слышим, как «общаются» люди, совершенно ничего не знающие о том, о чем они говорят; но когда человек, указывая на снежный ком, спрашивает «это тигр?», он ничего не знает о тиграх в совершенно ином смысле, если сравнивать с тем, в каком смысле человек, полагающий, что Ванкувер выиграет Кубок Стэнли или что война во Вьетнаме была начата с целью помочь южным вьетнамцам, ничего не знает о том, о чем он говорит. Очевидно, что в случае человека, думающего, что Ванкувер выиграет Кубок Стэнли или что война во Вьетнаме была начата с целью помочь южным вьетнамцам, мы не решим проблему, если примем определенные лингвистические конвенции; однако незнанию человеком того, о чем он говорит, во втором, ошеломляющем, смысле можно воспрепятствовать, — и практически это так и делается, — с помощью наших конвенций о языке. Я утверждаю, что говорящие *должны* знать хоть что-нибудь о (стереотипных) тиграх с тем, чтобы считаться усвоившими слово «тигр»; должны знать что-нибудь о вязах (или во всяком случае, об их стереотипе) с тем, чтобы считаться усвоившими слово «вяз» и т. д.

Эта идея не должна показаться слишком неожиданной. В конце концов, мы же не разрешаем людям вести машину на автострадах, если они не прошли определенных тестов, подтверждающих наличие у них некоторого *минимального* уровня знаний; мы не ужинаем с людьми, которые не научились пользоваться ножом и вилкой. Лингвистическое сообщество также имеет свой минимум стандартов как в отношении синтаксиса, так и в отношении «семантики».

Однако характер требуемого минимума знаний в значительной мере зависит от культуры и от рассматриваемой темы. В нашей культуре требуется, чтобы говорящие знали, как выглядят тигры (если они усвоили слово «тигр», что также считается обязатель-

ным); от них не требуется, чтобы они в деталях знали, как выглядит дерево вяз (например, какую форму имеют его листья). От людей, говорящих по-английски, их лингвистическое сообщество требует, чтобы они были способны отличить тигра от леопарда; но от них не требуется, чтобы они могли отличить вяз от бука.

Но вполне могло бы быть и иначе. Представьте себе индейское племя чероки, у которого есть, скажем, два слова: *ihaba'* и *wa'arabi* для обозначения вяза и бука соответственно, и для которого считается *обязательным* умение различать эти деревья. О черокезе, который не умел бы отличить вяз, сказали бы, что он не знает, что такое *ihaba'*, не знает значения слова *ihaba'* (или, возможно, не знает этого слова или не *владеет* им); точно так же об англичанине, не имеющем понятия о том, что тигры полосаты, сказали бы, что он не знает, что такое тигр, не знает значения слова «тигр» (если же он, по крайней мере, знает, что тигры — это большие кошки, мы могли бы сказать, что он знает часть значения или частично знает значение) и т. д. Тогда перевод слова *ihaba'* как «вяз», а слова *wa'arabi* — как «бук» был бы, с нашей точки зрения, только *приблизительно* правилен. В этом смысле радикальный перевод ¹⁹ представляет собой реальную трудность, но это не та абстрактная трудность, о которой говорит Куайн ²⁰.

ЧТО ТАКОЕ СТЕРЕОТИПЫ?

Я ввел понятие «стереотип» в лекциях, прочитанных мною в Университете Вашингтона и в Миннесотском Центре по философии науки в 1968 году. В опубликованной затем статье «Возможна ли семантика?» ²¹ эта тема развита дальше, а в настоящей статье я хотел бы вновь обсудить это понятие и ответить на некоторые возникшие в связи с ним вопросы.

В обыденном употреблении слово «стереотип» выражает общепринятое (часто предумышленное) представление (которое может быть совершенно неточным) о том, на что похож, чем является и

¹⁹ Термин введен Куайном (в книге «*Word and Object*»): он означает перевод, для которого не может стать подспорьем общая культура или родственные языки.

²⁰ Обсуждение вопроса о предполагаемой невозможности единственно правильного радикального перевода см.: Putnam H. *The Refutation of conventionalism* // Putnam H. *Mind, Language and Reality*. Philosophical Papers, Cambridge..., 1975, Vol. 2, p. 159—191.

²¹ Putnam H. *Is Semantics Possible?* // Putnam H. Указ. соч. p. 139—152.

как ведет себя *X*. Очевидно, что моя трактовка стереотипа включает некоторые черты обыденного употребления. Я не касаюсь предумышленных стереотипов (кроме тех случаев, когда сам язык таков); но меня интересуют общепринятые представления, которые могут быть неточными. Я думаю, что именно такое общепринятое представление связано со словами «тигр», «золото» и т. д., и, более того, это единственная доля истины в трактовке значения как «понятия».

С этой точки зрения тот, кто знает, что означает слово «тигр» (или, как мы решили говорить, — усвоил слово «тигр»), *должен* знать, что *стереотипные* тигры полосаты. Говоря точнее, лингвистическое сообщество, как таковое, требует от человека знание *одного* определенного стереотипа тигров (хотя человек может иметь и другие); от человека требуется, чтобы он знал этот стереотип и знал (имплицитно), что он обязателен. Этот стереотип должен включать признак полосатости, чтобы мы сочли человека усвоившим слово «тигр».

Включение какого-либо признака (например, полосатости) в стереотип, связываемый со словом «*X*», не означает, что аналитически истинно то, что все *X* обладают этим признаком, или что большинство *X* обладает этим признаком, или что все нормальные *X* обладают этим признаком, или что некоторые *X* обладают этим признаком²². Тигры с тремя ногами или тигры-альбиносы — это логически непротиворечивые существа. Если мы когда-нибудь установим, что наш стереотип основан на ненормальных или нетипичных представителях естественного вида, то это не будет обнаружением логического противоречия. Если бы тигры утратили свою полосатость, они от этого не перестали бы быть тиграми так же, как бабочки не перестали бы быть бабочками, если бы утратили свои крылья.

(Строго говоря, ситуация гораздо сложнее. Слову типа «бабочка» можно придать смысл, в каком бабочки, утратившие крылья, скажем вследствие мутации, перестанут быть бабочками. Таким образом, для слова «бабочка» можно найти смысл, в каком утверждение «бабочки имеют крылья» будет аналитической истиной. Думаю, однако, что более важным является тот смысл этого термина, в каком бабочки без крыльев не перестали бы считаться бабочками.)

Здесь у читателя может возникнуть вопрос — какую ценность представляют стереотипы для лингвистического сообщества, если «информация», содержащаяся в стереотипе необязательно является правильной. На самом деле это не такая уж и большая загадка.

²² Это обосновывается в статье: Putnam H. Is Semantics Possible? // Putnam H. Указ. соч. p. 139—152.

Большинство стереотипов действительно включают признаки, которыми обладают парадигмальные представители данного класса. Даже если стереотипы ошибочны, их ошибочность проливает свет на вклад, вносимый ими в коммуникацию. Стереотип золота, например, содержит признак *желтизны*, хотя химически чистое золото почти белое по цвету. Однако золото в ювелирных изделиях обычно бывает желтым (благодаря примеси меди), поэтому присутствие этого признака в стереотипе оказывается даже полезным в повседневных условиях. Стереотип, связываемый со словом *ведьма*, ошибочен в более серьезном отношении, по крайней мере, если речь идет о его экзистенциальной нагрузке. Если считать (с вытекающей экзистенциальной нагрузкой), что ведьмы находятся в сговоре с Сатаной и вызывают болезни и смерть и т. п., то этот стереотип облегчает коммуникацию только тогда, когда она не выходит за рамки учения о ведьмах. Это не облегчит коммуникацию в тех случаях, когда требуется не столько согласие с воззрениями других собеседников, сколько более адекватный взгляд на мир. (Строго говоря, я имею в виду стереотип, который существовал в Новой Англии триста лет назад; сегодня частью стереотипа о ведьмах является представление о том, что в *реальности* ведьм не существует, и тем самым сводятся на нет губительные последствия учения о ведьмах.) Но наличие в нашем языке некоторых стереотипов, которые скорее препятствуют, чем облегчают наше взаимодействие с миром и друг с другом, указывает только на то, что мы не являемся непогрешимыми, а иначе и быть не может. Между тем наша коммуникация вряд ли могла бы быть успешной, если бы большинство наших стереотипов не были настолько точными, насколько это возможно.

«ОПЕРАЦИОНАЛЬНОЕ ЗНАЧЕНИЕ» СТЕРЕОТИПОВ

Возникает один сложный вопрос: в какой мере понятие стереотип «операционально определимо». Здесь нужно быть предельно осторожным. Известно, что в физических науках провалились попытки дать *строго* операциональные определения понятий; а раз это не удалось в физике — нет оснований ожидать, что это удастся в лингвистике. Порой аргументы Куайна, доказывающие невозможность теории значения, сводятся к требованию операциональных определений в лингвистике; когда это имеет место, эти аргументы следует не принимать в расчет. Однако часто случается, что операциональные определения возможны не для реального мира, а для некоторой идеализированной ситуации. Формулировка таких «операциональных определений» имеет эвристическую ценность, как и любая другая

идеализация. Она становится опасной только тогда, когда мы по ошибке видим в операциональном определении нечто большее, чем удобную идеализацию. Поэтому мы можем спросить: каково «операциональное значение» утверждения о том, что слово имеет такой-то и такой-то стереотип, не предполагая, что ответ на этот вопрос будет теоретическим объяснением того, что значит быть стереотипом.

Для теоретического объяснения того, что значит быть стереотипом, необходимо понятие *лингвистического обязательства* — понятие, которое мы считаем фундаментальным для лингвистики и которое мы не будем пытаться здесь эксплицировать. Сказать, что полосатость — это часть (лингвистического) стереотипа слова «тигр», значит признать *обязательным* для человека, усвоившего слово «тигр», знать, что стереотипные тигры полосаты — в том же самом смысле «обязательности», в каком человек, говорящий о львах по-английски, обязан указывать, говорит ли он о них в единственном или во множественном числе. Нетрудно описать идеализированный эксперимент для проверки этой гипотезы. Представим себе человека, который выступает для лингвиста своеобразным *конфедератом* (*confederate*). Этот взрослый конфедерат в совершенстве владеет (или делает вид, что владеет) английским языком, однако по некоторым причинам (Вырос в чужой культуре? Воспитывался в монастыре?) он абсолютно не знает слова «тигр». Он произносит слово «тигр» или, еще лучше, указывает на это слово (как если бы он не знал, как его произнести) и спрашивает «что означает это слово?» или «что это?» и т. п. Если отвлечься от всех осложнений, связанных с проведением подобных экспериментов на практике, то по нашей гипотезе конфедерату в ответ должны, как правило, сказать, что тигры, *inter alia*, полосаты.

Вместо того, чтобы полагаться на конфедератов, лингвист мог бы исследовать, как учатся английскому языку дети. Однако детей, изучающих свой родной язык, учат не так часто, как полагают философы; детей не учат языку, они учатся ему сами, подчеркивал Хомский. И все же они иногда задают такие вопросы, как «что такое тигр?», и по нашей гипотезе в таких случаях в ответ им также должны сказать, *inter alia*, что тигры полосаты. Но проблема состоит в том, что скорее всего отвечают на эти вопросы родители, а темперамент родителей, время и внимание, которое они уделяют детям, сильно варьируют.

Было бы нетрудно привести множество других «операциональных» приложений нашей гипотезы, но в этом нет необходимости. Дело в том, что мы в полной мере владем знанием своего родного

языка и прекрасно понимаем свои лингвистические обязательства. Даже если представить дело так, будто мы находимся в позиции марсиан по отношению к английскому языку, то этим не будет достигнута методологическая ясность в этом вопросе; в конце концов, трансформационная лингвистика сформировалась в стройную науку только тогда, когда ученые отказались от операционального подхода.

Так, если бы кто-нибудь спросил у меня значение слова «тигр», то я прекрасно знаю, что бы я ответил. Я сказал бы, что тигры — это животные из семейства кошачьих, сказал бы что-нибудь об их размерах, о том, что они желтые с черными полосками, что иногда они живут в джунглях и что они свирепые животные. Я мог бы добавить и другие вещи — все зависело бы от того, в каком контексте и по какой причине был задан вопрос; но перечисленные позиции, за исключением, пожалуй, сведений о джунглях, я счел бы *обязательным* сообщить. Мне не нужно ставить эксперимент, чтобы установить, что именно это я считаю обязательным сообщить, и я уверен, что приблизительно такие же сведения сочтут обязательным сообщить и другие люди. Конечно, идиолекты изменяются от одного человека к другому; признак полосатости (если отвлечься от того, что считать «рисунком», а что — «фоном», т. е. видеть ли черные полоски на желтом фоне, как вижу их я, — или видеть желтые полоски на черном фоне) будет присутствовать во всех нормальных идиолектах; но информацию о том, что тигры (стереотипно) живут в джунглях, одни люди могли бы счесть обязательной, а другие — необязательной. Или же, по аналогии с определенными синтаксическими признаками одни характеристики стереотипа (быть большой кошкой, быть полосатым) можно было бы считать обязательными, а другие — *необязательными* (optional). Но мы не будем исследовать здесь эту возможность.

К ВОПРОСУ О «ДВУХ ДОГМАХ ЭМПИРИЗМА» КУАЙНА

В своей работе «Две догмы эмпиризма» Куайн предпринял мощную и плодотворную атаку на модное сейчас различие между аналитическим и синтетическим. Эта антитеза стала настоящим пожирателем философов: аналитическое *отождествлялось* с необходимым, необходимое *отождествлялось* с тем, что в принципе не подлежит пересмотру, а это последнее *отождествлялось* с любой истиной, какую отдельный философ не потрудился обосновать. Однако атака Куайна в некоторых аспектах зашла слишком далеко; на наш взгляд, можно сохранить некоторый ограниченный класс аналитических

предложений²³. Более важно то, что позже и сам Куайн, и другие философы сочли результатом этой атаки ниспровержение не только различия между аналитическим и синтетическим, но и всего понятия значения. Как уже отмечалось, мы согласны с тем, что традиционное понятие значения сталкивается с серьезными трудностями, но наша задача в данной статье не деструктивная, а конструктивная. Наша цель — пересмотреть понятие значения, а не похоронить его. Поэтому будет полезно рассмотреть, какие возражения против возможности пересмотра этого понятия содержатся в аргументах Куайна.

В целом, аргументы Куайна против понятия аналитичности сводятся к следующему: этому понятию нельзя приписать никакого поведенческого значения. Согласно его аргументу (в чем-то упрощающему ситуацию), имеется, по существу, два кандидата на роль поведенческого показателя аналитичности, и оба совершенно неудовлетворительны, правда, по разным причинам. Первый поведенческий показатель — *центральность* (centrality): многие современные философы называют предложение аналитическим, если некоторое сообщество (скажем, преподаватели Оксфорда) считает, что оно не подлежит пересмотру. Но, как убедительно показывает Куайн, максимальная неподверженность пересмотру не является исключительной прерогативой аналитических предложений. Предложения, выражающие фундаментальные законы физики (например, закон сохранения энергии), также обнаруживают в своем поведении максимальную неподверженность пересмотру, хотя непривычно и необоснованно определять эти законы как аналитические. Однако Куайн исходит не только из того, что необоснованно считать аналитическими все утверждения, от которых мы крайне неохотно отказываемся; он отмечает, что в реальной истории науки «неподверженность пересмотру» проявляется в *разной степени*. В реальной практике рациональной науки не существует такой вещи, как *абсолютная* неподверженность пересмотру. Таким образом, отождествление аналитичности с неподверженностью пересмотру изменило бы это понятие в двух отношениях: аналитичность стала бы вопросом степени и не стало бы абсолютно аналитических предложений. Это настолько расходилось с классическим — в духе Карнапа, Айера, *et al*, — понятием аналитичности, что Куайну стало понятно: если *это* — именно то, о чем мы собираемся говорить, то было бы меньше путаницы, введи мы для этого другое понятие, скажем, понятие *центральности*.

²³ См.: Putnam H. The Analytic and the Synthetic // Putnam H. Указ. соч., р. 33—69.

Второй поведенческий показатель — это *быть названным «аналитическим»*. По сути дела, для некоторых философов критерием аналитичности служит то, что предложение *называют* аналитическим специалисты (скажем, преподаватели Оксфорда). Этот показатель имеет следующие варианты: предложение должно выводиться из предложений некоторого конечного списка, в начале которого тот, кто унаследовал статус ученика Карнапа, начертил слова «Постулат Значения»; или предложение должно быть получено из теоремы логики путем подстановки синонимов на место синонимов. Последний из этих вариантов выглядит многообещающим, однако Куайн выдвигает против него вопрос «что считать критерием синонимии?». Такой критерий мог бы быть следующий: слова W_1 и W_2 — синонимы, если и только если эквиваленция (x) (x входит в экстенционал слова $W_1 \equiv x$ входит в экстенционал слова W_2) является аналитически истинной; но здесь мы попадаем в порочный круг. Согласно другому критерию, слова W_1 и W_2 синонимичны, если и только если их называют синонимами специалисты; но это все тот же второй показатель, только в слегка измененном виде. Перспективным выглядит и такой критерий: слова W_1 и W_2 синонимичны, если и только если они взаимозаменяемы *salva veritate* (т. е. их можно подставлять на место друг друга) в любых контекстах соответствующего класса. Однако Куайн убедительно показал, что и этот критерий ведет к порочному кругу. Таким образом, второй показатель сводится к следующему: предложение является аналитическим, если или оно само, или некоторое связанное с ним определенным способом выражение, или последовательность упорядоченных пар таких выражений, или множество таких выражений, принадлежат к классу, на всех членов которого специалисты ссылаются, произнося определенный звук: или звук АНАЛИТИЧЕСКОЕ, или звук ПОСТУЛАТ ЗНАЧЕНИЯ, или звук СИНОНИМИЧНОЕ. В конечном счете, при таком подходе «аналитическое» и т. п. остается *неэксплицированным звуком*.

Хотя Куайн не рассматривал этой возможности, однако ясно, что взяв пересечение классов предложений, выделяемых двумя неудовлетворительными поведенческими показателями, мы также не получим удовлетворительного результата; эксплицировать аналитичность предложения с помощью понятий «быть центральным» *плюс* «называться АНАЛИТИЧЕСКИМ» значит просто указать, что аналитические предложения образуют подкласс центральных предложений, не поясняя при этом, в чем состоит специфика этого подклас-

са. По сути дела, Куайн делает вывод, что аналитичность — это или ошибочно истолкованная центральность, или ничто.

Вопреки убедительным доводам Куайна, многие философы продолжают злоупотреблять понятием аналитичности, часто смешивая его с предполагаемой максимальной степенью центральности. Из альтернатив, указанных Куайном, они выбрали отождествления аналитичности с центральностью, и им пришлось в результате отнести такие синтетические по виду предложения как «пространство имеет три измерения» к разряду аналитических и признать существование в науке такой вещи, как абсолютная неподверженность пересмотру, несмотря на впечатляющие свидетельства в пользу обратного. Но этот подход можно опровергнуть, дополнив доводы Куайна важным аргументом Рейхенбаха.

Рейхенбах²⁴ показал, что есть некоторое множество принципов, каждый из которых Кант определил бы как синтетический *a priori*, но конъюнкция которых несовместима с положениями частной теории относительности и принципом общей ковариантности. (Это множество принципов включает обычную индукцию, непрерывность пространства и евклидово пространство.) Кантианец может последовательно держаться евклидовой геометрии, чего бы это ни стоило; однако последующий опыт может заставить его отказаться от обычной индукции или непрерывности пространства. Или же он может держаться обычной индукции и непрерывности пространства, чего бы это ни стоило, однако последующий опыт может заставить его отказаться от евклидовой геометрии (это имеет место тогда, когда физическое пространство даже не гомеоморфно какому-либо евклидову пространству). В другой статье²⁵ Рейхенбах формулирует по сути тот же самый аргумент, но в слегка измененном виде.

Применительно к нашему случаю, это доказывает, что существуют принципы, которые для философов, приверженных широкому понятию аналитичности и, в особенности, для философов, отождествляющих аналитичность с (максимальной) неподверженностью пересмотру, являются аналитическими, но конъюнкция которых имеет проверяемые эмпирические следствия. Стало быть, мы должны раз и навсегда отказаться от отождествления аналитичности с центральностью, иначе — мы вынуждены или отказаться от идеи, что конъюнкция сохраняет аналитичность, или же принять то неприятное

²⁴ Reichenbach H. The Theory of Relativity and A Priori Knowledge. California, 1965, p. 31.

²⁵ В кн.: Schilpp P. (ed.) Albert Einstein: Philosopher-Scientist. N. Y., 1951.

следствие, что аналитическое предложение может иметь проверяемые эмпирические следствия (и, следовательно, *аналитическое предложение может оказаться эмпирически ложным*).

Кстати, не случайно, что предложения, которые Кант определил бы как синтетические *a priori*, сегодняшние эмпиристы сочли бы аналитическими; раздувая понятие аналитичности, они как раз и стремились к тому, чтобы устранить кантовскую проблему, отождествив *априорность* с аналитичностью, а затем отождествив аналитичность с конвенциональной истиной. (Этот последний шаг Куайн также подверг уничтожительной критике, но обсуждение этого вопроса отвлекло бы нас от нашей темы.)

Другие философы попытались возразить Куайну, используя различие между *предложениями* и *утверждениями*: да, они согласны, что все *предложения* подвержены пересмотру, а вот некоторые *утверждения* — нет. Пересматривая предложение, мы не изменим нашего мнения относительно утверждения, прежде выражаемого этим предложением, если предложение (понимаемое как синтаксический объект вместе с его значением) после пересмотра фактически окажется не синонимичным тому же предложению до пересмотра, т. е. если при пересмотре изменится значение, а не теория. Но, во-первых, это сразу возвращает нас к попытке эксплицировать аналитичность в терминах синонимии; и, во-вторых, одна вещь представляет несомненный вклад Куайна в философию, это — осознание того, что нельзя четко отделить изменение значения от изменения теории. Мы не согласны с Куайном, что изменение значения вообще нельзя определить, но отсюда не следует обоснованность дихотомии «*или изменение значения, или изменение теории*». Открытие, что мы живем в неевклидовом мире, *могло бы изменить значение слова «прямая линия»* (это случилось бы в том — маловероятном — случае, если бы стереотип прямизны включал в себя что-то вроде постулата о параллельных прямых), но это не было бы *просто* изменением значения. В частности, это не было бы изменением *экстенционала*: так, было бы неправильно утверждать, что постулат о параллельных прямых был «истинным с учетом прежнего смысла слов». Из того факта, что отказ от предложения *S* предполагал бы изменение значения, не следует, что *S истинно*. Значения могут не соответствовать миру, а изменение значения может быть вызвано эмпирическими открытиями.

Хотя в нашей статье мы не пытаемся эксплицировать понятие аналитичности, но мы пытаемся эксплицировать понятие, которое может показаться тесно с ним связанным, т. е. понятие значения.

Поэтому может показаться, что аргументы Куайна направлены и против того, что пытаемся делать мы. Давайте разберемся в этом.

По нашему мнению, в одном, совершенно правильном, смысле полосатость образует часть значения слова «тигр». Но из этого не следует, с нашей точки зрения, что утверждение «тигры полосаты» является аналитическим. Могло бы случиться, что в результате мутации все тигры стали альбиносами. Коммуникация предполагает, что мой стереотип тигров включает их полосатость, и ваш стереотип включает полосатость, и вы знаете, что мой стереотип включает полосатость, и вы знаете, что я знаю... (и так далее, в манере *а-ля Грайс*, до бесконечности). Но *коммуникация* не предполагает, что любой конкретный стереотип правилен, или что большинство наших стереотипов уже навсегда правильны. Лингвистическая обязательность не является показателем неподверженности пересмотру или даже истинности; таким образом, мы можем считать утверждение «тигры полосаты» частью значения слова «тигр» и не быть при этом втянутыми в проблемы, связанные с аналитичностью.

Итак, аргументы Куайна против отождествления аналитичности с центральностью не направлены против отождествления входящих характеристик в состав «значения» слова *X* с ее обязательным включением в стереотип *X*. А аргумент Куайна о «звуках»?

Конечно, данные о том, что *говорят* люди, включая высказываемые ими металингвистические замечания, важны как для «семантики», так и для синтаксиса. Так, человек, указывающий на *моллюска* и спрашивающей «это тигр?», вызовет, видимо, у всех хохот. (Закончив хохотать), люди, возможно, скажут: «он не знает значения слова “тигр” или «он не знает, что такое тигры». Такие комментарии могут быть полезны лингвисту. Но мы не *определяем* стереотип с помощью подобных комментариев. Когда мы говорим, что быть «похожим-на-большую-кошку» образует часть значения слова «тигр», мы имеем в виду не только то, что отнесение слова «тигр» к чему-то, что не похоже-на-большую-кошку (и не является тигром) вызывает произнесение определенных *звуков*. Мы имеем в виду, что люди, усвоившие слово «тигр», знают, что «тигры (стереотипно) похожи на больших кошек», и считают своим долгом передать это знание тем, кого они учат употреблению этого слова. Минимум навыков, которых требует принадлежность к лингвистическому сообществу, представляет собой важную информацию, но *ни* какого круга, критикуемого Куайном, здесь не содержится.

РАДИКАЛЬНЫЙ ПЕРЕВОД

Наша теория не решает, во всяком случае, автоматически поставленную Куайном проблему «радикального перевода» (т. е. перевода с чужого языка или с языка, принадлежащего к чужой культуре). Мы не можем переводить с нашего гипотетического языка чероки на английский, просто соотнося стереотипы, поскольку установление стереотипа такого, скажем, слова как *wa'arabi*, предполагает перевод соответствующих выражений с языка чероки. С другой стороны, если установить требование, чтобы каждое слово в языке чероки — с учетом его стереотипа — отображалось на соответствующее английское слово (или приблизительно соответствующее, поскольку во многих случаях нельзя достичь точного соответствия), то это накладывает серьезные *ограничения* на переводную функцию (translation-function). Как только нам удастся перевести основную лексику в языке чероки, мы сможем приступить к выяснению стереотипов, а они помогут нам в будущем сузить класс возможных вариантов перевода и проверить правильность уже частично построенной переводной функции.

Даже если мы можем определить стереотипы (на основе, скажем, предварительного перевода «основной лексики»), их, как правило, недостаточно для выявления единственно возможного перевода. Так, немецкие слова «*Ulme*» и «*Buche*» имеют такой же стереотип, как и английское слово «*elm*» («вяз»); но «*Ulme*» означает «вяз», тогда как «*Buche*» означает «бук». При переводе с немецкого на английский однокоренной характер слов «*Ulme*» и «*elm*» мог бы указать, как правильно переводить «*Ulme*» (хотя это далеко не надежный способ, ибо, как правило, однокоренные слова не являются синонимами); а в случае греческого языка у нас нет такого ключа и нам ничто не подсказывает, какое из слов «*ὄξυα*» и «*πτελέα*» означает *вяз*, а какое — *бук*; мы должны были бы найти грека, который смог бы отличить *вяз* от *бука* (или *oxya* от *ptelea*). Это показывает, что, возможно, не установление *типичной* предрасположенности носителей языка к согласию или несогласию должно заботить лингвиста, ибо из-за разделения лингвистического труда часто лингвист должен вначале определить, кто выступает экспертами в отношении таких слов, как *oxya*, *wa'arabi*, *gavagai* и т. п. И только потом он сможет высказывать догадку о социально определяемом экстенционале этих слов. В последующем *и* социально определяемый экстенционал, *и* стереотип *типичного* носителя языка будут накладывать ограничения на переводную функцию. Если мы установим, что стереотип

охуа очень отличается от стереотипа «вяза», то это исключит перевод *охуа* как «вяза» во всех контекстах, кроме наиболее экстенциональных; а если мы установим, что класс вязов даже приблизительно не является *экстенсионалом* слова *охуа*, то это исключит указанный вариант перевода во всех контекстах.

Заметьте, что мы расширили совокупность фактов, учитываемых при определении переводной функции, выйдя за рамки того аскетического набора, который допускает Куайн в своей книге «Слово и объект». Например, Куайн не допускает «знание» лингвистом того, что люди говорят то-то и то-то, когда «конфедерат» лингвиста указывает на слово *охуа* и спрашивает «что оно означает?» или «что это?», — не допускает на том основании, что этот род «знания» предполагает, что уже имеется перевод вопроса «что означает это слово?». Однако, если Куайн готов допустить, что можно *каким-то образом* угадать слова, выражающие согласие или несогласие в чужом языке, то нет ничего неразумного и в предположении о том, что можно каким-то образом сообщить туземцу о своем непонимании какого-то слова. Для этого вовсе не нужно точно знать, как на чужом языке формулируется вопрос: «Что означает это слово?» (в противовес таким фразам, как: «Я не понимаю это слово», «Это слово мне не знакомо» или «Я озадачен этим словом» и т. п.). Возможно, будет достаточно просто произнести с недоумением слово *охуа* или любое другое слово. Почему *недоумение* должно быть менее доступно понимаю лингвиста, чем *согласие*?

Кроме того, мы можем использовать то обстоятельство, что сегментация на слова присуща всем языкам (имеются даже тесты по определению сегментации на слова и морфемы без знания значения). Ясно, что если мы позволяем лингвисту произносить целые предложения, чтобы установить согласие или несогласие, то у нас нет и серьезных оснований запрещать ему произносить слова и морфемы недоуменным тоном.

Повторяю, я не *выдвигаю* здесь требования, что такое расширение набора свидетельств разрешает проблему радикального перевода. Оно только накладывает дальнейшие ограничения на класс допустимых вариантов перевода. Уверен, что, расширяя класс ограничений, можно определить единственно возможный перевод, насколько, конечно, это достижимо в реальной практике. Однако, на мой взгляд, следует использовать и ограничения, которые не имеют непосредственного отношения к лингвистической теории; так, следует учитывать, какого рода верования правомерно приписывать людям

(каковы связи верований друг с другом, с культурой и внешним миром). Обсуждение этих вопросов мы отложим до другого раза.

КРИТИКА СЕМАНТИЧЕСКОЙ ТЕОРИИ ДЭВИДСОНА

В ряде публикаций Дональд Дэвидсон выдвинул интересную идею — сформулировать семантическую теорию естественного языка, взяв за образец то, что в математической логике называется *определением истины* для формализованного языка. Если не вдаваться в технические подробности, то эта идея сводится к следующему: формулируется множество правил, определяющих (1) для каждого слова при каких условиях оно истинно относительно какого-либо объекта (имеются в виду слова, к которым применимо понятие экстенционала; все другие слова следует рассматривать как синкатегорематические); а (2) для предложений, содержащих более одного слова формулируются правила, определяющие условия истинности каждого предложения в зависимости от способа построения этого предложения из более коротких предложений (слова рассматриваются как если бы они были односложными предложениями, например, слово «снег» считается предложением «это снег»). Выбор односложных предложений в качестве отправной точки — это моя интерпретация проекта Дэвидсона; но в любом случае он предполагал начать с *конечного* множества *коротких* предложений, для которых *непосредственно* определены условия истинности. Условие (2) не означает, что должно существовать правило для каждого предложения, не охватываемого условием (1), поскольку это потребовало бы бесконечного множества правил; оно означает, что должно существовать правило для каждого *типа* предложения. Например, в формализованном языке к этому типу правил можно отнести правило: если S есть $(S_1 \ \& \ S_2)$ для некоторых предложений S_1 и S_2 , то S истинно, если и только если и S_1 , и S_2 истинны.

Отметим, что в этом примере условие истинности, сформулированное для предложений типа $(S_1 \ \& \ S_2)$, задает значение «&». Точнее, оно определяет значение структуры $(— \ \& \ —)$. Именно в этом смысле определение истины может быть теорией значения. По мнению Дэвидсона, в такой форме можно задать всю теорию значения для естественного языка.

Несомненно, правила только что описанного типа могут определять значение некоторых слов и структур. Вопрос в том, какие у нас есть основания считать, что значение большинства слов можно определить таким образом, не говоря уже обо всех словах.

Проблема очевидна: для многих слов можно построить экстенционально правильное определение истины, которое ни в коей мере не будет теорией значения этих слов. Например, рассмотрим определение: «Слово “вода” истинно относительно x , если и только если x есть H_2O ». Это экстенционально правильное определение истины для слова «вода» (строго говоря, это не определение истины, а определение «истинности», или *выполнимости* в смысле Тарского, но нас не будут беспокоить здесь такие нюансы). Это определение экстенционально правильно, по крайней мере, в том случае, если мы не учитываем, что воду с примесями также называют «водой» и т. п. Допустим, большинство людей не знают, что вода — это H_2O . Тогда эта формула ничего не говорит нам о значении слова «вода». Она может представлять интерес для химика, но ее нельзя считать теорией значения для термина «вода». Ее можно считать теорией экстенционала для термина «вода», но Дэвидсон обещает нам гораздо большее.

Дэвидсон очень хорошо осознает эту трудность. Его ответ (во всяком случае, в беседе) состоит в том, что нам нужно разработать теорию перевода. Подобно Куайну, он считает это реальной проблемой. Формулируемая с учетом теории перевода (которой мы, надо признаться, еще не располагаем), теория значения сводится к следующему: нам нужна система определений истины, которая одновременно была бы системой переводов (или приблизительных переводов, если совершенный перевод недостижим). Если бы у нас была теория, устанавливающая, что считать хорошим переводом, то мы могли бы исключить приведенное выше определение истины для слова «вода» как не представляющее интереса на том основании, что выражение « x есть H_2O » — неприемлемый перевод или даже вовсе не перевод выражения « x есть вода» (для ненаучного сообщества), хотя истинно, что вода = H_2O .

Отсюда уже совсем недалеко до утверждения, что теория значения — это определение истины плюс теория значения. (Если бы у нас была яичница с беконом, она была бы у нас, *будь* у нас бекон и *будь* у нас яйца.) Но, как мы увидим дальше, худший недостаток этой истории вовсе не в том, что она только обещает.

Согласно второму тезису Дэвидсона, в теории перевода, которой у нас пока нет, исходными элементами должны быть *предложения*, а не слова, поскольку в качестве *свидетельств* в лингвистике используются только факты согласия или несогласия с предложениями. Слова же, утверждает Дэвидсон, можно трактовать как предложения (слово «вода» — как предложение «это вода» и т. д.)

Как можно оценить с позиции подхода, развиваемого в настоящей статье, этот амбициозный проект построения теории значения в форме определения истины с ограничениями теорией перевода, для которой предрасположенность носителей языка к использованию предложений предоставляет «единственно доступные свидетельства»?

Наш ответ таков: этот проект неосуществим в принципе. В особых случаях, например, когда рассматривается слово «и» в его истинностно-функциональном смысле, определение истины (строго говоря, один из шагов в том, что логики называют «определением истины», ибо совокупность всех этих шагов и образует индуктивное определение истины для конкретного языка) может задать значение слова или структуры, поскольку стереотип, связываемый со словом (если имеет смысл говорить о стереотипе в случае такого слова, как «и»), действительно образует необходимое и достаточное условие. Если бы все слова были такими, как слова «и» и «холостяк», проект можно было бы выполнить. Безусловно, Дэвидсон внес важный вклад, подчеркнув, что лингвистика должна использовать индуктивно определяемые условия истинности. Но для большинства слов требования к теории истины и требования к теории значения взаимно несовместимы, по крайней мере в случае, когда мы строим теорию значения для английских слов и формулируем ее на английском языке. Но именно этот случай является основным.

Проблема состоит в том, что, как правило, со словом X совпадают по экстенционалу и имеют приблизительно одинаковый стереотип только выражения, сами содержащие слово X . Если мы исключим такие определения истины (строго говоря, шаги в определении, но я, для простоты, буду и дальше использовать выражение «определение истины» как для отдельных шагов, так и для всей их совокупности), как

Предложение « X есть вода» истинно, если и только если X есть вода

на том основании, что они ничего не говорят о значении слова «вода», и если мы исключим такие определения истины, как

Предложение « X есть вода» истинно, если и только если X есть H_2O .

на том основании, что они неправильно описывают значение слова «вода», то мы останемся ни с чем.

Проблема состоит в том, что нам нужно такое определение

Слово W истинно относительно x , если и только если «—»

которое удовлетворяло бы следующим условиям: (1) оно должно быть экстенционально правильным («—» означает условие, содержащее « x », например, « x есть H_2O »); (2) условие «—» должно быть, согласно нашей теории, переводом слова W , а это означает, что стереотип, связываемый со словом W , должен приблизительно совпадать со стереотипом, связываемым с «—»; (3) условие «—» не должно содержать ни самого слова W , ни его синтаксических модификаций. Если, к примеру, мы берем в качестве W слово «вяз», то невозможно выполнить все эти три условия одновременно. В любом экстенционально правильном определении условие «—», содержащее слова «вяз», будет абсолютно непригодным в качестве перевода слова «вяз».

Даже если в языке имеется два точных синонима, ситуация не-намного лучше. Так, определение

Слово «беладонна» истинно относительно x , если и только если x есть красавка

верно, равно как верно и определение

Слово «красавка» истинно относительно x , если и только если x есть беладонна

но разве это можно считать теорией значения для слов «красавка» и «беладонна»?

Заметьте, что именно условие (3) логики не считают нужным соблюдать в своих определениях истины. Определение истины

Предложение «Снег бел» истинно, если и только если снег бел

служит образцом для логиков. Однако их цель — определить не значение предложения «снег бел», а экстенционал слова «истинный» применительно к конкретному языку. Тарский рискнул бы даже сказать, что он определил значение (а не просто экстенционал) слова «истинный»; но он никогда не стал бы утверждать, что его определение истины говорит что-то о значении предложения «снег бел».

Возможно, в действительности Дэвидсон думает, что теория значения в любом серьезном смысле этого слова невозможна, и мы можем лишь сформулировать переводные функции. Если это так, то он вполне мог бы считать, что единственно возможная «теория значения» для английского языка — это утверждать, что «слово “вяз” истинно относительно x , если и только если x есть вяз», «слово “вода” истинно относительно x , если и только если x есть вода» и т. д., и хотя изредка встречаются такие случаи, как « S_1 & S_2 истинно, если и только если и S_1 и S_2 истинно». Однако если «теория» Дэвидсона представляет собой лишь скептицизм Куайна под маской позитивного вклада в исследование значения, то это горькая пилюля.

Более того, считать, что единственные, доступные свидетельства лингвист черпает из предрасположенности носителей языка к использованию целых предложений, значит придерживаться точки зрения, которая при одной интерпретации оказывается пустой, а при другой — совершенно ложной. Если предрасположенность говорить определенные вещи в ответ на *вопрос об отдельных словах, морфемах или синтаксических структурах* включается в понятие предрасположенности использовать предложения, то принимая только предрасположенность к использованию предложений, мы, видимо, ничего не исключаем. При другой интерпретации, точка зрения Дэвидсона, по сути, сводится к тому, что лингвисту не доступны данные, сообщаемые носителями языка (в том числе и самим лингвистом), когда их спрашивают о значении слова, морфемы или синтаксической структуры. Однако еще не было указано ни одной причины, почему лингвисту не доступны такие данные, но совершенно очевидно, что реальные лингвисты во многом опираются на свидетельства носителей языка, когда изучают чужой язык, и на свою собственную интуицию, когда изучают родной язык. Непонятно, почему при переводе целого предложения мы не можем руководствоваться нашим знанием о синтаксических и семантических свойствах составных частей этого предложения, включая данные о его глубинной структуре. Как мы видели, есть методика сбора информации об отдельных составных частях предложения. Примечательно, что методика, которую Куайн и Дэвидсон считают единственно *возможной* и которая основана на первичности целых предложений, а не отдельных слов, идет *вразрез* с тем, чему мы обязаны всеми достигнутыми успехами в изучении естественного языка.

КРИТИКА КАЛИФОРНИЙСКОЙ СЕМАНТИКИ

Теперь я хотел бы рассмотреть подход к семантической теории, впервые сформулированный поздним Карнапом. Чтобы не заниматься разбором текстов, я не стану приписывать указанную точку зрения какому-либо конкретному философу, а буду просто говорить о «калифорнийской семантике».

Введем понятие *возможного мира*. Пусть f — функция, определяемая на «пространстве» всех возможных миров, и ее значением $f(x)$ в любом возможном мире x выступает подмножество множества сущностей в мире x . Назовем f *интенционалом*. Термин T имеет значение для человека X , если X связывает с термином T интенционал f_T . Термин T *истинен* относительно сущности e в возможном мире x , если и только если e принадлежит ко множеству $f(x)$. Карнап был склонен говорить об «усвоении» интенционалов, а не о «связывании» интенционалов с терминами; однако, несомненно, для него X не просто «усваивает» интенционал f , он усваивает f как интенционал T , т. е. он некоторым образом *связывает* f с T .

Совершенно очевидно, что такая трактовка понимания термина расходится с тем, что предлагается в настоящей статье. В ответ сторонник калифорнийской семантики мог бы сказать, что его семантика дает описание *идеального языка*; а реальный язык *неопределенен*. Иными словами, в реальном языке термин T не имеет одного точно определенного интенционала; он имеет множество — возможно, нечеткое множество — интенционалов. Однако первый шаг на пути описания естественного языка, несомненно, связан с изучением той идеализации, при которой каждый термин T имеет строго один интенционал.

(В своей книге «*Значение и необходимость*» Карнап использует другую формулировку: интенционал — это просто *свойство*. Сущность e входит в экстенционал термина T только в том случае, если e обладает тем свойством, которое составляет интенционал термина T . Предыдущая формулировка с использованием функции f позволяет не рассматривать *свойство* как исходное понятие.)

Первая трудность, с которой сталкивается такая точка зрения, связана с использованием совершенно неразъясненного понятия *усвоения* интенционала (или, в нашей переформулировке, *связывания* интенционала с термином). Определив интенционалы как теоретико-множественные сущности f , мы получаем «конкретную» математическую реализацию этого понятия в современном стиле (на основе понятий возможного мира и множества), но в результате нельзя по-

нять, как можно иметь интенционал в сознании, что значит думать об интенционале, «усваивать» интенционал или «связывать» его с чем бы то ни было. Здесь не поможет указание на то, что думать об интенционале значит использовать слово или функциональный заместитель слова (будь то аналог слова в «мозговом коде», если признать, что мозг «вычисляет» с помощью «кода», который имеет сходство с языком и, возможно, что-то заимствует из языка; или будь то такая форма мысли, как образ или личный символ, — в тех случаях, когда они используются в мышлении) для референции на данный интенционал, поскольку сама референция (т. е. вхождение объекта в экстенционал термина) была определена нами с помощью интенционала. Хотя вполне можно подобным образом описывать, что значит думать о такой абстрактной сущности, как функция или свойство, в настоящем контексте это описание явно содержит в себе порочный круг. И никакого не содержащего круга описания для этого фундаментального понятия еще не было предложено.

Эта трудность связана с более общей проблемой в философии математики, отмеченной Полем Бенасеррафом²⁶. Согласно Бенасеррафу, разные теории в философии математики постоянно оказываются между двух стульев: или им удастся объяснить природу математических объектов и необходимость математической истины, но не удастся объяснить тот факт, что люди могут обучаться математике, сослаться на математические объекты и т. п., или же они могут объяснить второе, но не могут объяснить первое. Калифорнийская семантика может объяснить, что такое интенционал, но не может избежать порочного круга, объясняя, как мы «усваиваем» интенционалы, связываем их с терминами, думаем о них, *ссылаемся на них* и т. д.

Возможно, Карнап не заметил эту трудность из-за своего верификационизма. Первое время он считал, что понимать термин значит *уметь верифицировать*, входит ли любая данная сущность в экстенционал термина или нет. В этом случае «усвоение» интенционала означало бы умение верифицировать, принадлежит ли сущность e в любом возможном мире x к $f(x)$ или нет. Позже Карнап изменил свою точку зрения, осознав, что предложения, говоря языком Куайна, оказываются перед судом опыта сообща, а не по отдельности. Не существует такого способа верифицировать истинность термина T относительно некоторой сущности, который не за-

²⁶ Benacerraf P. Mathematical Truth // Journal of Philosophy, 8 November 1973 (70:19, 661—679).

висел бы от контекста, составленного из множества теорий, вспомогательных гипотез и т. д. Не исключено, что Карнап мог бы считать какой-то вариант этой ранней теории верным в случае ограниченно-го класса терминов — так называемых «терминов наблюдения». С нашей точки зрения, верификационная теория значения ложна как в своей основной идее, так и в применении к терминам наблюдения, но мы не будем это обсуждать здесь. Во всяком случае, тому, кто не является верификационистом, трудно видеть в калифорнийской семантике теорию, поскольку понятие *усвоения* интенционала остается в ней совершенно необъясненным.

Во-вторых, если допустить, что «усвоение интенционала» (связывание интенционала с термином *T*) — это *психологическое состояние* (в узком смысле), то калифорнийская семантика принимает оба принципа (1) и (2), которые мы подвергли критике в первой части настоящей статьи. Она должна признать, что психологическое состояние носителя языка определяет интенционал его терминов, который в свою очередь определяет их экстенционал. Следовательно, если два человека находятся в одном и том же психологическом состоянии, то они с необходимостью приписывают один и тот же экстенционал любому используемому ими термину. Как мы видели, это совершенно неверно в отношении естественного языка. Причина этого, как мы установили выше, отчасти состоит в том, что экстенционал определяется социальными факторами, а не одним только индивидуальным знанием языка. Таким образом, калифорнийская семантика вынуждена рассматривать язык как нечто индивидуальное и полностью игнорировать разделение лингвистического труда. Экстенционал любого термина рассматривается в этой школе как полностью определяемый тем, что находится в голове индивидуального человека. Вторая причина, почему это неверно, как мы видели, состоит в том, что большинство терминов являются *жесткими десигнаторами*. В калифорнийской семантике каждый термин рассматривается, по сути дела, как *дескрипция*. Игнорируется индексальный компонент значения, т. е. тот факт, что наши термины обозначают объекты, сходные, в некоторых отношениях, с объектами, которые мы обозначаем *жестко*, — сходные с *этими* объектами, с веществом, которое мы *здесь* называем «водой» и т. п.

А как быть с утверждением о том, что сторонник калифорнийской семантики имеет дело не с реальным языком, а с идеализацией, которая «не учитывает неопределенность», и что термины естественного языка можно считать связанными скорее со множеством интенционалов, чем с одним точно определенным интенционалом?

В ответ укажем, что *индексальное* слово нельзя представить как неопределенное семейство *неиндексальных* слов. Слово «я», если взять крайний случай, является *индексальным*, но не *неопределенным*. «Я» не синонимично какой-либо *дескрипции*; не синонимично оно и неопределенному множеству дескрипций. Аналогично, если мы правы, то слово «вода» не синонимично ни дескрипции, ни неопределенному множеству дескрипций (интенционалов).

Аналогичным образом, слово, экстенционал которого фиксируется социальнo, а не индивидуальнo, и слово, экстенционал которого фиксируется *неопределенно* и индивидуальнo, — это разные вещи. Причина, почему «усвоенный» мной смысл слова «вяз» не фиксирует экстенционал вяза, не в том, что слово является неопределенным — если бы проблема состояла просто в неопределенности, то из совпадения моих понятий вяза и бука следовало бы, что в моем употреблении вяза — это буки или крайние случаи буков, а буки — это вяза или крайние случаи вязов. Причина же состоит в том, что экстенционал слова «вяз» в моем диалекте фиксируется не тем, что средний носитель языка «усваивает» или же не «усваивает»; он фиксируется сообществом, включающим экспертов, в ходе сложного процесса сотрудничества. Язык, воплощающий в себе разделение лингвистического труда, даже приблизительно нельзя смоделировать как язык, содержащий неопределенные термины и не предполагающий лингвистическое разделение труда. Сотрудничество не есть неопределенность.

Однако можно было бы возразить: разве нельзя заменить наш реальный язык языком, в котором (1) термины были бы заменены на *неиндексальные* термины с такими же экстенционалами (например, слово «вода» заменено на слово « H_2O », при условии, что слово « H_2O » *неиндексально*); и (2) разделение лингвистического труда исчезло бы в силу того, что каждый стал бы экспертом по любому вопросу?

Мы должны ответить отрицательно на этот вопрос; но представим на минутку, что ответ утвердительный. Какое бы это имело значение? Полученный «идеальный» язык ни в коей мере не стал бы подобием нашего реального языка, и дело здесь отнюдь не в «неопределенности естественного языка».

Однако эту замену невозможно осуществить по той причине, что *все* термины естественных видов и термины физических величин индексальны в описанном нами смысле — слово «водород» и, стало быть, слово « H_2O » так же индексальны, как и слово «вода». Возможно, термины «чувственных данных», если таковые имеются, *неиндексальны* (исключая такие термины, как «я»), но слово «жел-

тый» как предикат *вещного* языка индексально по той же самой причине, что и слово «тигр»: даже если нечто *выглядит* желтым, оно может *не быть* желтым. Если нам возразят, что вещи, которые выглядят желтыми при нормальных обстоятельствах (и для людей с нормальным зрением), являются желтыми, то это не поможет, так как слово «нормальный» предполагает здесь именно ту характеристику, которую мы называем индексальностью. Нет никаких оснований считать, что проект сведения нашего языка к неиндексальному языку в принципе может быть осуществлен.

Думаю, что устранить разделение лингвистического труда «в принципе» можно. Однако, если разделение лингвистического труда присуще, как я догадываюсь, всем языкам, то какой интерес представляет возможность языка, не обладающего основополагающей характеристикой *человеческого* языка? В мире, где каждый выступает экспертом по любому вопросу, социальные законы невероятно отличаются от тех, какие мы имеем сейчас. *На каком основании* мы должны принимать такой мир и такой язык в качестве модели для анализа *человеческого* языка?

Кстати, философы, работающие в традиции калифорнийской семантики, недавно попытались устранить именно эти недостатки в своей модели. Так, было выдвинуто предположение, что интенционал — это функция, аргументами которой являются не только возможные миры, а, вероятно, возможный мир, носитель языка и нелингвистический контекст, в котором произносится слово. Это позволило бы представить в модели некоторые виды индексальности и некоторые виды разделения лингвистического труда. Согласно Дэвиду Льюису, разрабатывающему эти идеи, слово «вода», например, имело бы один и тот же интенционал (одну и ту же функцию) на Земле и на Двойнике Земли, но разные экстенционалы. (По сути дела, Льюис сохраняет допущение (1), сформулированное в первой части настоящей статьи, и отбрасывает допущение (2); мы же отбрасываем допущение (1) и сохраняем (2).) Вполне возможно, что формальные модели, разработанные Карнапом и его последователями, окажутся ценными при такой модификации. Нас же интересовала здесь не полезность математического формализма, а философия языка, лежащая в основе ранних вариантов этого подхода.

СЕМАНТИЧЕСКИЕ МАРКЕРЫ

Если предложенный здесь подход верен, то предстоит проделать большую работу с тем, чтобы (1) определить, какого рода компонен-

ты могут входить в стереотипы; (2) выработать подходящую систему представления стереотипов и т. д. Однако эту работу нельзя выполнить в ходе философского обсуждения. Скорее, это входит в компетенцию лингвистики и психолингвистики. И здесь, я думаю, может представлять ценность идея *семантического маркера*. Эта идея была сформулирована в работах Дж. Каца и Дж. Э. Фодора, и мы воспользуемся ею в несколько модифицированном виде.

Рассмотрим стереотип слова «тигр». Он включает такие признаки как «быть животным»; «быть похожим-на-большую-кошку», «иметь черные полосы на желтом фоне» (или «желтые полосы на черном фоне») и т. д. В признаке *животное* есть что-то особенное. С точки зрения куайновского понятия *центральности* или *неподверженности пересмотру*, этот признак качественно отличается от других перечисленных признаков. Конечно, мы можем представить, что тигры могли бы не быть животными (они могли бы быть роботами). Но разясняя этот случай дальше, мы видим, что тогда они должны были быть роботами всегда; мы вовсе не имеем в виду тигров, *замененных* роботами, поскольку тогда роботы не были бы тиграми. Если же они не всегда были роботами, то они должны были бы *стать* роботами, а это еще трудней вообразить. Если тигры есть и всегда были роботами, то эти роботы, должно быть, не очень «разумны», или иначе не могло бы быть так, что тигры не есть животные — мы, скорее, описывали бы тот случай, когда некоторые роботы являются животными. Лучше всего было бы представить их роботами, «контролируемыми другими» — скажем, оператором на Марсе, управляющим на расстоянии каждым их движением. Повторяю, очень трудно дать здесь разяснения, и удивительно трудно придумать, с чего начать; именно поэтому легко сделать ошибочный вывод, что для тигра «логически невозможно» *не* быть животным. С другой стороны, нетрудно представить отдельного тигра без полос; он может быть альбиносом. Нетрудно представить конкретного тигра, который не выглядит как большая кошка: он может быть чудовищно изуродован. Мы можем даже представить разновидность тигров, утративших свои полоски или чудовищно изуродованных. Но тигры, переставшие быть животными? Опять очень трудно представить!

Отметим, что мы не совершаем ошибку, которую критиковал Куайн, и не приписываем абсолютную неподверженность пересмотру таким утверждениям, как «тигры — это животные», «тигры не могли бы превратиться из животных во что-то другое и при этом остаться тиграми». Действительно, мы можем описать неестественные ситуации, в которых такие утверждения были бы отброшены

как ложные. Но мы считаем, что *качественно* труднее пересмотреть предложение «все тигры — животные», чем предложение «все тигры полосаты» — по сути, последнее утверждение можно и не считать истинным.

Такие признаки, как «животное», «живое существо», «создаваемый человеком предмет», «день недели», «период времени» не только как признаки с высокой центральностью связываются со словами «тигр», «моллюск», «вторник», «час», но и образуют часть широко используемой и важной *системы классификации*. Их центральность гарантирует, что объекты, отнесенные к этим рубриками, фактически никогда не придется классифицировать *заново*, поэтому эти рубрики используются в качестве естественных категориальных указателей (category-indicator) во множестве контекстов. Мне представляется разумным по аналогии с синтаксисом, где мы используем при классификации такие маркеры, как «имя существительное», «прилагательное», а в более узкой области, «конкретное имя существительное», «глагол, требующий одушевленного подлежащего и абстрактного дополнения» и т. д., использовать в семантике в качестве маркеров эти категориальные указатели.

Примечательно, что, вводя идею семантического маркера, Кац и Фодор не предполагали, что значение (в нашей терминологии, стереотип) исчерпывается перечислением таких маркеров. Скорее, в число маркеров включались только категориальные указатели высокой центральности, что предлагаем сделать и мы. Остальные признаки просто перечислялись как входящие в состав «отличителя». Схему этих авторов трудно сравнивать с нашей, ибо им хотелось бы, чтобы семантические маркеры *плюс* отличитель всегда обеспечивали необходимое и достаточное условие вхождения объекта в экстенционал термина. Поскольку предполагалось, что маркеры и отличитель в совокупности составляет то, что каждый человек знает имплицитно, они должны были признать, что каждый человек имплицитно знает необходимое и достаточное условие вхождения объекта в экстенционал таких терминов, как «золото», «алюминий», «вяз», а это, как мы уже отмечали, не всегда так. Затем Кац выдвинул требование, что *все* признаки составляют *аналитически* необходимое и достаточное условие вхождения в экстенционал, и ему пришлось отбросить различие между маркерами и отличителями; если все признаки обладают центральностью, так сказать, в бесконечной степени, то почему одни из них нужно называть «маркерами», а другие — «отличителями»? С нашей точки зрения, первоначальное различие между «маркерами» и «отличителями» бы-

ло правильным — если опустить идею, что отличитель обеспечивает (вместе с маркерами) необходимое и достаточное условие, и что в этом и состоит теория *аналитической* истины. На наш взгляд, идея семантического маркера, как мы ее понимаем, представляет собой важный вклад в теорию значения.

ЗНАЧЕНИЕ «ЗНАЧЕНИЯ»

Теперь мы можем подвести итог сказанному и предложить, каким образом можно было бы переформулировать понятие «значения». Предлагаемое нами решение — не единственно возможное, но оно поможет выделить некоторые важнейшие моменты. Кроме того, я думаю, оно максимально сохраняет то из обыденного употребления слова «значение», что, видимо, можно и уместно сохранить. Но поскольку, на мой взгляд, в том или ином виде допущения (I) и (II), сформулированные в первой части настоящей статьи, глубоко коренятся в обыденном представлении о значении, а эти допущения, взятые вместе, противоречат фактам, то невозможно переформулировать понятие «значение» без некоторых несогласующихся с интуитивным представлением следствий.

Говоря коротко, я предлагаю определить «значение» не путем выбора некоторого объекта, который затем отождествляется со значением (хотя, если потребуется, это можно сделать в обычной теоретико-множественной манере), а путем задания нормальной формы (*normal form*) (или, скорее, *типа* нормальной формы) для описания значения. Если мы знаем, каким должно быть «описание нормальной формы» для значения слова, то, как я понимаю, мы знаем, *что* такое значение в любом представляющем интерес для науки смысле.

Предлагаемое мной решение состоит в том, что описание нормальной формы для значения слова представляет собой последовательность или «вектор», который должен содержать следующие компоненты (возможно, уместно включить и другие типы компонентов): (1) синтаксические маркеры, применяемые к данному слову, например, «имя существительное»; (2) семантические маркеры, применяемые к данному слову, например, «животное», «период времени»; (3) дескрипция дополнительных признаков стереотипа, если таковые имеются; (4) дескрипция экстенционала.

Мое решение включает следующую конвенцию: все компоненты вектора, *исключая экстенционал*, выражают гипотезу относительно знания языка отдельным человеком. Так, описание нормальной формы для слова «вода» могло бы быть таким:

СИНТАКСИЧЕСКИЕ УКАЗАТЕЛИ

собирательное имя существительное, конкретное

СЕМАНТИЧЕСКИЕ УКАЗАТЕЛИ

естественный вид; жидкость

СТЕРЕОТИП

бесцветная, прозрачная, без вкуса, утоляющая жажду и т. д.

ЭКСТЕНСИОНАЛ

H₂O (может содержать примеси)

но это не означает, что знание о том, что вода есть H₂O, вменяется отдельному носителю языка или даже обществу. Это *мы* утверждаем, что экстенсионалом термина «вода», как употребляют этот термин *они* (предполагаемые носители языка), *на самом деле* является H₂O. Возражение: «кто *мы* такие, чтобы утверждать, что на самом деле является экстенсионалом *их* термина» — обсуждалось выше. Отметим, что, в основном, это возражение касается понятия *истины*, а экстенсионал, как связанное с истиной понятие, наследует ее проблемы.

Назовем два описания значения *эквивалентными*, если они во всем одинаковы, за исключением дескрипции экстенсионала, хотя экстенсионал в этих двух описаниях один и тот же. Стало быть, если экстенсионалом данного слова *действительно* является множество, имеющее разные дескрипции в двух описаниях значения, а другие компоненты описания значения правильно представляют различные аспекты знания языка, то оба описания являются правильными. Эквивалентные описания или одновременно правильны, или одновременно неправильны. Это иной способ выразить ту мысль, что, хотя мы должны использовать *дескрипцию* экстенсионала, чтобы *задать* экстенсионал, мы считаем, что в качестве компонента в описание входит *экстенсионал (множество)*, а не дескрипция экстенсионала.

В частности, слова «вода» в диалекте Земли и диалекте Двойника Земли будут описаны одинаково, за тем исключением, что в последней колонке описание нормальной формы для слова «вода» в диалекте Двойника Земли будет XYZ, а не H₂O. Это означает, с учетом сказанного, что мы приписываем одно и то же лингвистическое знание обычным жителям Земли и жителям Двойника Земли, но вместе с тем — разный экстенсионал слову «вода».

Предлагаемое нами означает, что мы сохраняем допущение (II), о котором речь шла ранее. Значение определяет экстенционал — в силу, так сказать, своей структуры. Однако допущение (I) отбрасывается; психологическое состояние конкретного носителя языка не определяет, «какое значение имеют его слова».

Думаю, в большинстве контекстов это согласуется с тем, как мы используем язык. Но вот один парадокс: допустим, Оскар свободно говорит на двух языках — немецком и английском. С нашей точки зрения, в рамках всей совокупности его диалектов слова «*beech*» и «*Vische*»²⁷ являются *точными синонимами*. Описания нормальной формы для их значения будут тождественны. Но Оскар может и не знать, что эти слова — синонимы! Носитель языка может иметь в своем лексиконе два синонима и не знать, что они синонимы!

Полезно рассмотреть, как в нашем анализе связаны между собой неправильность такого очевидного, на первый взгляд, рассуждения, как «если S_1 и S_2 синонимы, и Оскар понимает и S_1 , и S_2 , то Оскар знает, что S_1 и S_2 являются синонимами» и ложность допущения (I). Отметим, что если бы мы решили исключить экстенционал из компонентов «вектора значения», что, как я понимаю, предлагает сделать Дэвид Льюис, то мы столкнулись бы с тем парадоксом, что слова «вяз» и «бук» имеют *одно и то же значение*, но разные экстенционалы!

Как утверждает почти любая материалистическая теория, полагать что-то (верить в некоторое высказывание), значит осуществлять обработку (processing) некоторой *репрезентации* этого высказывания, будь то предложение в языке, часть «мозгового кода», форма мысли и т. п. Материалисты, и не только они, не склонны думать, что в мышлении можно иметь дело с высказываниями *в чистом виде*. Но даже материалисты, как правило, считают, что высказывания, репрезентации *которых* используются в мышлении, являются (простите за каламбур) нематериальными. Пусть и S_1 , и S_2 — две *имеющиеся* у меня репрезентации: если я верю в высказывание, выраженное с помощью репрезентации S_1 , я также должен верить в это высказывание, когда оно выражено с помощью репрезентации S_2 , — по крайней мере, я должен делать это, если хочу быть рациональным. Но это, как мы только что видели, неверно. Оскар вполне может считать, что *это* — «*beech*» (это несет на себе знак, говорящий «*beech*»), но не считать, что это — «*Vische*». Дело не только в том, что считать что-то — это процесс, предполагающий

²⁷ Оба переводятся как «бук».

репрезентацию; Оскар верит в высказывание (если вообще нужно предполагать «высказывания») при одной его репрезентации и не верит — при другой.

В теории значения изумляет то, как долго этот предмет находился в тисках неверных философских представлений, и как сильны эти неверные представления. Один за другим философы отождествляли значение с необходимым и достаточным условием. В эмпиристской традиции один за другим философы отождествляли значение с методом верификации. Эти две позиции отнюдь не исключают друг друга; не так уж мало было философов, которые считали, что значение = методу верификации = необходимому и достаточному условию.

С другой стороны, удивляет то, насколько слабым оказалось воздействие фактов. В конце концов, в настоящей статье отмечены не более, чем обыденные истины о том, как мы употребляем слова и как много (или, скорее, как мало) мы на самом деле знаем, когда их употребляем. Сам я стал размышлять над этими вопросами после того, как опубликовал статью, в которой с уверенностью утверждал, что значением слова является «совокупность семантических правил»²⁸, а затем засомневался, как можно таким образом объяснить значение обычного слова «золото». Нельзя сказать, чтобы философы никогда не обращались к подобным случаям: Локк, например, приводит это слово в качестве примера, но у него не возникает никаких сомнений в том, что значение этого слова является необходимым и достаточным условием!

Если и есть причина, почему мнение и ученых, и непрофессионалов оказалось столь ошибочным в отношении предметов, с которыми каждый сталкивается в повседневном опыте, о которых у всех нас имеется больше данных, чем мы можем осознать, и есть, если отбросить предвзятые мнения, весьма ясные интуиции, — то эта причина должна быть связана с тем, что существовавшие всегда и существующие теперь невероятно ошибочные воззрения о языке отражают две характерные и очень важные тенденции в философии: тенденцию трактовать познание как чисто *индивидуальный* процесс и тенденцию игнорировать *мир* в той мере, в какой он составляет нечто большее, чем совокупность индивидуальных «наблюдений». Игнорировать разделение лингвистического труда значит игнорировать социальный аспект познания; игнорировать то, что мы называли *индексальностью*, значит игнорировать вклад, вносимый окружаю-

²⁸ Putnam H. How not to talk about meaning // Putnam H. Указ. соч., р. 117—131.

шей средой. Традиционная философия языка, подобно традиционной философии, не принимает во внимание существование других людей и мира; более адекватная философия и более адекватная наука о языке должны включить их в область своего исследования.