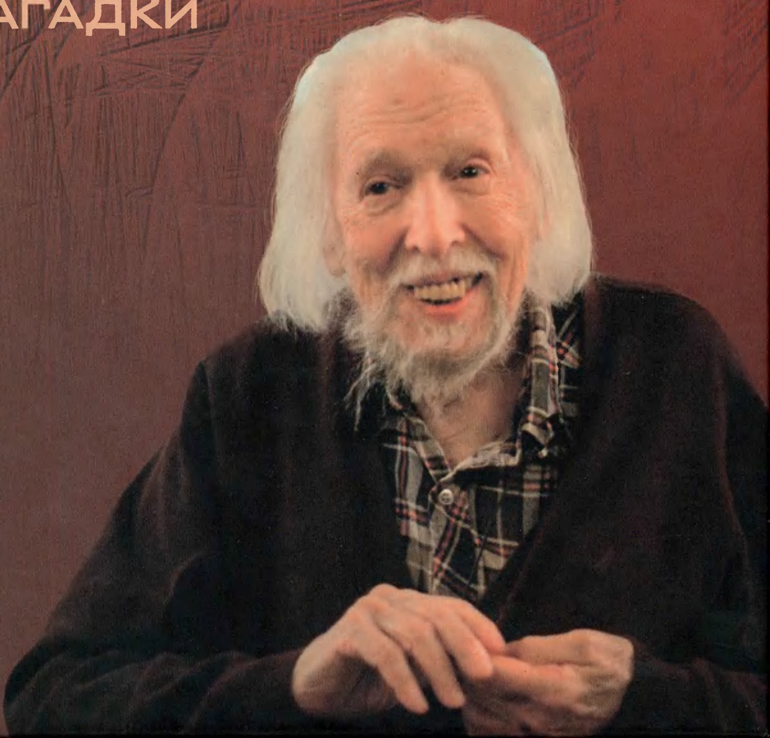
Библиотека аналитической философии

АФ

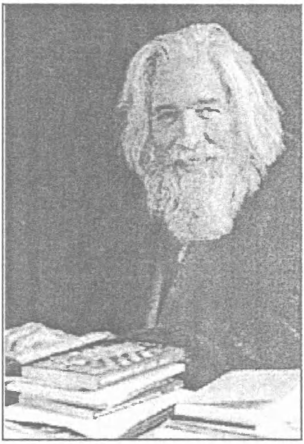
Раймонд Смаллиан

Вовеки неразрешимое

Путь к Геделю через занимательные загадки



БИБЛИОТЕКА
АНАЛИТИЧЕСКОЙ
ФИЛОСОФИИ



Раймонд Смаллиан

Раймонд СМАЛЛИАН

ВОВЕКИ НЕРАЗРЕШИМОЕ

Путь к Геделю через занимательные загадки

By Raymond SMALLYAN

FOREVER UNDECIDED

A Puzzle Guide to Godel

Oxford University Press, 2000

МОСКВА

КАНОН+

2013

УДК 1/14
ББК 87.6
С50

Смаллиан Раймонд

С50 Вовеки неразрешимое. Путь к Геделю через занимательные загадки / Раймонд Смаллиан. Пер. В.В. Целищева. — М.: «Канон$^+$» РООИ «Реабилитация», 2013. — 303 с.

ISBN 978-5-88373-081-7

Эта книга представляет собой введение в теоремы Геделя посредством логических занимательных проблем с применением математической логики. Аргументация Геделя перенесена из формальной области математических систем в область идей, более доступных обычному читателю. Основной упор сделан на системы вер и их соотношению с математикой. Это приводит к семантике возможных миров, которая играет существенную роль в компьютерных исследованиях и искусственном интеллекте.

УДК 1/14
ББК 87.6

ISBN 978-5-88373-081-7

Посвящается
всем непротиворечивым
мыслителям,
которые никогда
не смогут узнать,
что они непротиворечивы

ПРЕДИСЛОВИЕ

Возможно ли для рационального человека находиться в таком состоянии, когда он не может поверить в свою собственную непротиворечивость, не теряя в то же время этой самой непротиворечивости? Это одна из главных тем данной книги. Она смоделирована на знаменитом открытии Курта Геделя (так называемая Вторая Теорема о Неполноте), что любая достаточно сильная непротиворечивая математическая система, известная как элементарная арифметика, подвержена удивительному ограничению, суть которого в том, что она не может доказать своей собственной непротиворечивости!

Есть несколько причин, по которым я перевел аргумент Геделя из формальной области математических систем и доказуемых предложений в область человеческих существ и предложений, в которые они *верят*. Прежде всего, люди с их верами гораздо более понятны неспециалистам, нежели абстрактные математические системы, и поэтому я могу объяснить суть идей Геделя на языке, понятном любому человеку. Кроме того, изложение этих вопросов в терминах общечеловеческих представлений имеет огромную психологическую привлекательность и оказывается в высшей степени близким к быстро развивающимся исследованиям в области искусственного интеллекта.

Как и в своих предшествующих книгах с логическими загадками, я начинаю с массы проблем о лгунах и правдолюбах (Плуты и Рыцари). Помимо новизны (почти все они публикуются здесь впервые), есть еще одна особенность – а именно то, что книга представляет собой введение в символическую логику (предмет *символической логики* преподается сейчас во многих университетах), испещренное логическими загадками. В книге объясняется, как это ценная логическая техника может систематически помочь в разрешении логических загадок о Рыцарях и Плутах. (Это должно представлять интерес для преподавателей курса логики разного уровня – для тех, кто желает оживить эти курсы творческими и занимательными задачами). Следовательно, знания, которые приобретет читатель в области элементарной логики, помогут ему в прояснении более трудных материй в работах Геделя в последующих главах книги.

Другая тема в этих главах – которая представляет интерес не в меньшей степени и для искусственного интеллекта – это так называемые *самовыполнимые веры*. При каких условиях простая вера в предложение может сделать эту веру истинной? (Может ли это иметь нечто общее с религией?) Существует важная теорема логика М.Г. Лёба, которая в высшей степени уместна в этих вопросах. Эта теорема близко соотносится с теоремой Геделя и, будучи сформулирована сперва в терминах Мыслителей и их вер, легко доступна рядовому читателю.

Хотя мой основной упор был на системах вер, это ни в коем случае не означает, что я останавливаюсь на этом. По мере развертывания материала книги, читатель приобретает все более глубокое понимание того, как эти системы вер соотносятся с важными системами математики. Это, в свою очередь, ведет нас к философски восхитительному предмету *семантики возможных миров* – области исследований, открытой Лейбницем и доведенной до совершенства логиком Солом Крипке. Этот предмет также играет важную роль в настоящее время в компьютерных исследованиях и искусственном интеллекте.

Я преподавал довольно долго материал, который изложен в книге, самым различным группам студентов и аспирантов, изучавших математику, компьютерные науки и философию. Реакция этих студентов была вознаграждением для меня – они были заинтригованы. В самом деле, любой неофит, который проявляет смекалку в математике или естественных науках, может полностью освоить эту книгу (хотя это и потребует некоторого старания). Но и многие эксперты найдут здесь весьма новый и свежий материал.

Просто замечательно, как логика, философия, психология, искусственный интеллект и компьютерные исследования, а также математика сближаются все больше в наши дни. Мы живем в восхитительное время!

Раймонд Смаллиан
1986 г.

ЧАСТЬ I

ВЫ МОГЛИ БЫ УДИВИТЬСЯ!

ГЛАВА 1. *Дьявольская загадка*

Я верю, что предлагаемая ниже загадка является, быть может, самой дьявольской из всех существующих загадок (и если это так, я гордо заявляю свои права на ее изобретение).

Два человека – А и В – делают друг другу предложения, которые приведены ниже. Проблема заключается в том, чтобы определить, чье предложение лучше.

А предлагает: Ты делаешь утверждение. Если утверждение истинно, ты получаешь десять долларов. Если утверждение ложно, тогда ты получаешь меньше десяти или больше десяти долларов, но не точно десять.

В предлагает: Ты делаешь утверждение. Независимо от того, является ли утверждение истинным или ложным, ты получаешь больше десяти долларов.

Какое из предложений вы предпочтете? Большая часть решает, что предложение B лучше, так как оно *гарантирует* больше десяти долларов, в то время как в предложении A нет определенности в выигрыше более десяти долларов. И на самом деле кажется, что предложение B лучше, но видимость зачастую обманчива. А теперь я делаю свое предложение: если некоторые из вас желают сделать мне предложение A, я заплачу двадцать долларов заранее. Кто-нибудь готов играть? (Перед тем как принять решение, вам лучше прочитать оставшуюся часть главы!)

Я не приведу решения этой проблемы, пока сначала не рассмотрим более простые, родственные задачи.

В своей книге *To Mock Mockingbird* я представил следующую загадку: Предположим, я предлагаю два приза – Приз 1 и Приз 2. Вы должны сделать утверждение. Если утверждение истинно, тогда я должен дать вам один из двух призов (не говоря, какой именно). Если ваше утверждение ложно, тогда вы не получаете приза. Ясно, что вы можете быть уверены в выигрыше одного из призов, делая следующее утверждение: «Два плюс два равно четыре», но предположим, что вы нацелились на Приз 1; какое утверждение вы должны сделать, чтобы гарантировать выигрыш Приза 1?

Мое решение таково. Вы говорите: «Ты не дашь мне Приз 2». Если это утверждение ложно, тогда оно говорит, что не тот случай, который описан в утверждении, и значит, что я *дам* вам Приз 2. Но я не могу дать вам приз за ложное утверждение, и поэтому утверждение не может быть ложным. Следовательно, оно должно быть истинным. Так как оно истинно, то, что оно говорит, имеет место, что означает, что вы не получите Приз 2. Но так как ваше утверждение истинно, я должен отдать вам один из призов, и так как это не Приз 2, это должен быть Приз 1.

Как мы вскоре увидим, эта маленькая загадка имеет близкое отношение в знаменитой Теореме о неполноте Геделя. Для того чтобы понять, какое именно отношение, рассмотрим похожую загадку (на самом деле, это та же загадка, только под другой формулировкой). Прежде всего, мы прибываем на Остров Рыцарей и Плутов (который играет важную роль в этой книге), на котором каждый Обитатель есть либо Рыцарь, либо Плут. На этом острове Рыцари могут делать только истинные утверждения, а Плуты – только ложные.

Предположим, что на этом острове есть два клуба – Клуб I и Клуб II. Только Рыцарям позволяется быть членами любого из клубов, а Плутам строго запрещено быть членами любого из них. Также каждый Рыцарь есть член одного и только одного из двух клубов. Вы однажды посещаете остров и встречаете незнакомого Обитателя, который делает утверждение, из которого вы делаете вывод, что он должен быть членом Клуба I. Какое утверждение сделал Обитатель, чтобы из него можно было сделать такой вывод?

По аналогии с последней проблемой, говорящий мог бы произнести вот что: «Я не член Клуба II». Если бы говорящий был Плутом, тогда он не мог бы быть членом Клуба II, и сделал бы при этом истинное утверждение, чего Плут сделать не мог. Следовательно, он должен быть Рыцарем и его утверждение должно быть истинным; он на самом деле не является членом Клуба II. Но так как он Рыцарь, он должен быть членом клуба – поэтому он принадлежит Клубу I.

Аналогия должна быть ясной: Клуб I соответствует тем, кто делает истинные утверждения и получает Приз 1. Клуб II соответствует тем, кто делает истинные утверждения и получает Приз 2.

Эти загадки воплощают в себе существенные идеи, лежащие в основе знаменитого геделева предложения, которое утверждает свою собственную недоказуемость в данной математической системе. Предположим, что мы классифицируем все *истинные* предложения системы (подобно Рыцарям в загадке) на две группы: Группа I состоит из всех предложений системы, которые, хотя и истинны, не *доказуемы* в системе. Группа II состоит из всех предложений, которые не только истинны, но на самом деле доказуемы в системе. Гедель сумел сконструировать предложение, которое утверждает, что оно было в Группе II, – предложение может быть перефразировано, «Я не доказуемо в системе». Если предложение ложно, тогда то, что оно утверждает, не будет иметь места, что должно означать, что оно *доказуемо* в системе, чего не может быть (так как все предложения, доказуемые в системе, истинны). Отсюда, предложение должно быть истинным, и, как оно утверждает, не доказуемо в системе. Таким образом, геделево предложение истинно, но не доказуемо в системе.

Мы еще много чего скажем о геделевых предложениях в ходе этой книги. А сейчас я хочу рассмотреть некоторые варианты Загадки о Призах.

1. Первый вариант

Я опять предлагаю два приза – Приз 1 и Приз 2. Если вы сделаете истинное утверждение, я дам вам, по крайней мере, один из призов, а возможно, и оба. Если вы сделаете ложное утверждение, вы не получите приза. Предположим, что вы стремитесь получить оба приза. Какое бы утверждение вы сделали?

Решение: Утверждение, которое сработает, таково: «Я либо выиграю оба приза, либо ни одного». Если это утверждение ложно, тогда он говорит, что это не имеет места, что означает, что вы получите точно один приз. Но вы не можете получить приз за ложное утверждение. Следовательно, утверждение должно быть истинным, и вы на самом деле либо получаете оба приза, либо не получаете ни одного. Так как вы не сделали ложного утверждения, которое не принесет вам приза, вы должны получить оба приза.

2. Второй вариант

В этот раз правила немного изменяются. Если вы сделали истинное утверждение, вы получаете Приз 2; если вы сделали ложное утверждение, вы не получаете Приз 2 (вы могли получить Приз 1 или не получить его). Какое утверждение выиграет Приз 1?

Решение: Просто говорите: «Я не получу приз». Если утверждение истинно, тогда, с одной стороны, вы не получаете приз (как говорит это утверждение), а с другой стороны, вы получаете Приз 2 за то, что сделали истинное утверждение. Это есть явное противоречие, отсюда, утверждение должно быть ложным. Тогда, в отличие от того, что утверждение говорит, вы должны получить приз. Вы не можете получить приз за ложное утверждение, поэтому вы получаете Приз 1.

3. Извращенный вариант

Вообразим, что в извращенном состоянии ума я скажу вам, что если сделаете ложное утверждение, вы получите один из призов, но если вы сделаете истинное утверждение, вы не получите приз. Какое утверждение выигрывает Приз 1?

Решение: В этот раз это такое утверждение: «Я выиграю Приз 2». Я оставляю доказательство этого читателю.

4. Больше о дьявольской загадке

Теперь вернемся к загадке в начале этой главы. Дьявольское в ней было как раз *мое* предложение заплатить двадцать долларов за то, что я сделал предложение А, потому что, если бы вы приняли мое предложение, я мог бы выиграть у вас сколько угодно – скажем, миллион долларов. Можете ли вы видеть, почему это так?

Решение: Все, что я должен сделать, это сказать: «Ты никогда не заплатишь мне точно ни десять долларов, ни миллион долларов». Если мое утверждение истинно, тогда, с одной стороны, вы не платите мне ни десяти долларов, ни миллиона долларов, но с другой стороны, вы должны заплатить мне точно десять долларов за истинное утверждение. Это противоречие. Отсюда, это утверждение не может быть истинным, и должно быть ложным. Так как оно ложно, оно говорит, что это не имеет места, что означает, что вы платите мне либо в точности десять долларов, или в точности миллион долларов, но вы не можете заплатить мне точно десять долларов за ложное утверждение, отсюда, вы должны заплатить мне миллион долларов.

Ну, что, продолжаем играть?

ГЛАВА 2. *Удивлены?*

Теперь обратимся к парадоксу экзамена-сюрприза – парадоксу, который имеет непосредственное отношение к открытиям Геделя, которые мы будем рассматривать в данной книге. Мы рассмотрим его в следующей форме: Утром в понедельник профессор говорит своему классу: «Я проведу экзамен на этой неделе, который будет для вас сюрпризом. Он может состояться сегодня, во вторник, в среду, четверг или, самое позднее, в пятницу. Утром в день экзамена, когда вы придете в класс, вы не будете знать, что это день экзамена».

Ну, студент-логик размышляет следующим образом: «Ясно, что экзамена не будет в последний день, в пятницу, потому что, если экзамена не будет в четверг, тогда я буду знать, что он будет в пятницу, и экзамен не будет сюрпризом. Это исключает пятницу, и поэтому я знаю, что последним возможным днем будет четверг. И если экзамена не будет в среду, тогда я буду знать, что экзамен будет в четверг, поскольку пятница уже исключена, но тогда сюрприза опять не получится. Так что четверг также исключается».

Таким же образом студент исключает среду, затем вторник и, наконец, понедельник, тот самый день, когда профессор обещает сюрприз. Тогда студент заключает: «Следовательно, экзамена не будет вообще; профессор не может выполнить свое обещание». А профессор говорит немедленно: «А сейчас экзамен». Студент удивлен в высшей степени!

В чем неправильность размышления студента?

Об этой знаменитой проблеме написана масса статей, и все же общего согласия по поводу ее разрешения не достигнуто. Моя точка зрения вкратце такова.

Пусть я оказался на месте студента. Я утверждаю, что у меня будет экзамен-сюрприз в любой день, *даже в пятницу*! Вот мой ход размышления: Предположим, что наступает утро пятницы, а экзамена все еще нет. Во что я должен тогда верить? Предполагая, что я верю, прежде всего, профессору (и это предположение необходимо для решения проблемы), мог бы я непротиворечиво продолжать верить профессору утром в пятницу, если бы экзамена все еще не было? Я не вижу такой возможности. Я мог бы определенно верить, что экзамен будет сегодня (в пятницу), но не мог бы верить, что это будет для меня *сюрпризом*. Следовательно, как я могу доверять точности в утверждениях профессора? Имея сомнения в отношении профессора, я просто не знал бы, чему верить. Все могло бы случиться, насколько это касается меня, и поэтому могло бы быть так, что я *мог бы* быть удивлен, обнаружив, что в пятницу будет экзамен.

На самом деле, профессор сказал две вещи: (1) У вас будет экзамен в один из дней недели; (2) Вы не знаете утром того дня, когда будет экзамен, что он состоится. Я полагаю важным отделить эти два утверждения. Вполне *могло бы* быть, что профессор прав в отношении первого утверждения и неправ в отношении второго. Утром в пятницу я не мог бы непротиворечиво верить, что профессор прав в отношении обоих утверждений, но я вполне непротиворечиво мог бы верить его первому утверждению. Однако, если бы я сделал это, его второе утверждение оказалось бы неверным (так как я буду верить тогда, что у меня *будет* экзамен сегодня). С другой стороны, если я сомневаюсь в первом утверждении профессора, я не знаю, будет или нет сегодня экзамен, что означает, что второе утверждение профессора выполняется (в предположении, что он сдержит свое слово и назначит экзамен). Поэтому удивительно, что второе утверждение профессора истинно или ложно в зависимости от того, верю я или не верю в его первое утверждение. Таким образом, единственный вариант, при котором профессор может быть прав, состоит в том, что у меня есть сомнения относительно него; если я сомневаюсь в нем, это делает его правым, в то время как если я доверяю ему, это делает его неправым! Я не знаю, была ли эта любопытная точка зрения принята во внимание в предыдущих дискуссиях по этому поводу.

Версия одного дня. Запутывание ситуации с несколькими днями на самом деле не имеет отношения к сути проблемы, поскольку есть «однодневная» ее версия. Профессор говорит студенту: «У тебя будет сегодня экзамен-сюрприз». Что студент может вывести из этого?

Эквивалентная проблема такова: Предположим, что студент спрашивает профессора теологии: «Существует ли Бог?» Профессор отвечает ему: «Бог существует, но ты никогда не поверишь, что Бог существует». Какой вывод может сделать студент из этого? Мы увидим позднее, что при определенных вполне разумных предположениях о мыслительных способностях студента, он не сможет поверить профессору, не впадая при этом в противоречие.

Проблема, имеющая еще большее отношение к теме этой книги, такова: Опять-таки студент спрашивает профессора, существует ли Бог. В этот раз профессор дает следующий любопытный ответ: «Бог существует, если и только если ты никогда не поверишь, что Бог существует». Эквивалентно профессор мог бы сказать: «Если Бог существует, тогда ты никогда не поверишь, что Бог существует, но если Бог не существует, тогда ты поверишь, что Бог все-таки существует». Что может заключить студент из *этого*? Может ли студент поверить профессору, не впадая при этом в противоречие? Да, оказывается, может, но (опять-таки при определенных вполне разумных предположениях о мыслительных способностях студента, суть которых будет объяснена позднее в этой книге) студент, который верит профессору, может оставаться непротиворечивым, если только он не *знает*, что он непротиворечив! Другими словами, если студент верит профессору, а также верит в свою собственную непротиворечивость, тогда он становится противоречивым.

Этот парадокс близко соотносится со знаменитой Второй Теоремой о Неполноте Геделя – теоремой о недоказуемости непротиворечивости. Гедель рассмотрел некоторые из наиболее всеобъемлющих математических систем. Эти системы определенно непротиворечивы (так как все доказанное в них является истинным), но удивительно то, что эти системы, несмотря на их силу (или же, если посмотреть на это по-другому, из-за их силы), неспособны доказать свою собственную непротиворечивость. Мы знаем о непротиворечивости этих систем при использовании только таких методов, которые не могут быть формализованы в самих этих системах.

В этой книге будут исследованы такие парадоксы и результаты Геделя. Для того чтобы облегчить нам это предприятие (и развлечься также!), давайте сначала обратимся к некоторым логическим загадкам и рассмотрим их соотношение с пропозициональной логикой.

ЧАСТЬ II

ЛОГИКА ИСТИНЫ И ЛЖИ

ГЛАВА 3. *Собиратель налогов*

Многие события в этой книге будут происходить на Острове Рыцарей и Плутов, где, как мы видели, Рыцари всегда делают истинные утверждения, а Плуты – всегда ложные и каждый Обитатель Острова является либо Рыцарем, либо Плутом.

Фундаментальный факт об этом Острове состоит в том, что невозможно для Обитателя заявить о том, что он является Плутом, потому что Рыцарь никогда не лжет и не говорит, что он Плут, а Плут никогда не допустит правды, а именно того, что он Плут.

Следующие четыре проблемы вводят логические связки *и, или, если-то, если-и-только-если*, которые будут рассмотрены более формально в главе 6.

Визит МакГрегора

Собиратель налогов МакГрегор однажды был по службе на острове Рыцарей и Плутов. На этом острове женщины также называются Рыцарями и Плутами. МакГрегор решил, что будет опрашивать только семейные пары.

1. (И)

МакГрегор постучал в одну дверь; муж приоткрыл ее и спросил, что тому нужно. «Я сборщик налогов, – ответил МакГрегор, – и мне нужна информация о вас и вашей жене. Кто из вас может быть Рыцарем, или кто из вас может быть Плутом?»

«Мы оба Плуты!», – сказал сердито муж и захлопнул дверь.

К какому типу принадлежит муж и к какому типу принадлежит жена?

Решение: Если бы муж был Рыцарем, он никогда не сказал бы, что он и его жена оба Плуты. Следовательно, он должен быть Плутом. Так как он Плут, его утверждение ложно; поэтому неверно, что они оба Плуты. Это означает, что его жена должна быть Рыцарем. Следовательно, он – Плут, а она – Рыцарь.

2. (Или)

В следующем доме МакГрегор спросил мужа: «Вы оба Плуты?» Муж ответил: «По крайней мере, один из нас Плут».

К каким типам принадлежат муж и жена?

Решение: Если бы муж был Плутом, тогда было бы истинным, что, по крайней мере, один из них Плут, отсюда, Плут должен был сделать истинное утверждение, чего не может быть. Следовательно, муж должен быть Рыцарем. Отсюда следует, что его утверждение было истинным, что означает, что-либо он, либо его жена является Плутом. Так как он не Плут, Плутом является его жена. И поэтому ответ противоположен Проблеме 1: он – Рыцарь, а жена – Плут.

Следующая проблема является более озадачивающей, чем предыдущие две (по крайней мере, тем, кто не сталкивался с подобными проблемами). Она содержит тему, которая проходит через все более сложные проблемы, которые разбросаны в последующих главах.

3. (Если, то)

Визит МакГрегора в следующий дом оказался более озадачивающим. Дверь открыл застенчивый человек. После того как МакГрегор попросил его дать сведения о себе и его жене, все, что муж произнес, было следующее: «Если я – Рыцарь, тогда моя жена – тоже Рыцарь».

МакГрегор ушел не очень довольный собой. «Как я могу сказать что-нибудь о каждом из них при таком неопределенном ответе?» – подумал он. Он был готов уже написать: «Муж и жена неизвестного типа», когда внезапно вспомнил старую лекцию по логике в дни студенчества. «Конечно, – понял он, – я могу сказать, к каким типам они *оба* принадлежат!»

К какому типу принадлежит муж, и к какому типу принадлежит жена?

Решение: Предположим, что муж – Рыцарь. Тогда истинно, что он сказал, а именно: что если он – Рыцарь, тогда его жена – тоже Рыцарь. Это доказывает, что *если* он – Рыцарь, Рыцарем является и его жена. Ну, это как раз в точности то, что сказал муж; он сказал, что *если* он Рыцарь, то Рыцарем является его жена. Следовательно, он сделал истинное утверждение и поэтому он должен быть Рыцарем. Мы теперь знаем, что он – Рыцарь, и мы уже доказали, что если он – Рыцарь, то Рыцарем является и его жена. Следовательно, муж и жена являются Рыцарями.

Идея, которая лежит в основе последней проблемы, имеет далеко идущие более тонкие следствия, чем может подумать читатель. Давайте рассмотрим следующий вариант проблемы: Предположим, что вы посещаете остров, перспективный с точки зрения наличия на нем золота. Перед тем как начать копать, вы хотите узнать, есть ли золото на самом деле на острове. Предполагается, что каждый Обитатель острова знает, есть на острове золото или нет. Предположим, что Обитатель говорит вам: «Если я – Рыцарь, тогда на острове есть золото». Вы можете тогда обоснованно заключить, что Обитатель должен быть Рыцарем, и золото должно быть на острове. Ход размышления таков же, как и в Проблеме 3: Предположим, что Обитатель есть Рыцарь, тогда на самом деле истинно, что если он Рыцарь, тогда золото есть на острове, и отсюда, золото есть на острове. Это доказывает, что если он – Рыцарь, тогда золото есть на острове. Так как он сказал именно это, он есть Рыцарь. Отсюда, золото есть на острове.

Решение Проблемы 3 и ее вариантов есть специальный случай следующего факта, который достаточно важен, чтобы сформулировать его в виде Теоремы I.

Теорема I. Дано некоторое предложение p, предположим, что Обитатель острова говорит: «Если я – Рыцарь, тогда p». Тогда Обитатель должен быть Рыцарем, а p должно быть истинно.

Решение Проблемы 3 есть специальный случай Теоремы I, где в качестве предложения p выступает предложение, что жена Обитателя есть Рыцарь. Вариант Проблемы 3 (о золоте на острове) есть также специальный случай Теоремы I, если в качестве p взять предложение, что на острове есть золото.

Из Теоремы I также следует, что никакой Обитатель острова Рыцарей и Плутов не может сказать: «Если я Рыцарь, тогда существует Санта Клаус» (если, конечно, Санта Клаус на самом деле не существует).

4. (Если и только если)

Когда сборщик налогов опрашивал четвертую пару, муж сказал: «Моя жена и я одного типа; мы либо оба Рыцари, либо оба Плуты».

(Муж мог альтернативно сказать: «Я есть Рыцарь, если и только если, моя жена есть Рыцарь». Это одно и то же.)

Что можно вывести относительно типа мужа и типа жены?

Решение: Нельзя определить, к какому типу принадлежит муж, но тип жены определяется следующим образом.

Если бы жена была Плутом, муж не мог бы утверждать, что он того же типа, что и жена, потому что это было бы равносильно утверждению, что он – Плут, утверждению, которого он не мог сделать.

Альтернативный путь рассмотрения этой проблемы таков: Муж является либо Рыцарем, либо Плутом. Если он – Рыцарь, его утверждение истинно, отсюда, он и его жена принадлежат одному типу, что означает, что его жена тоже Рыцарь. С другой стороны, если муж – Плут, его утверждение ложно, отсюда, он и его жена принадлежат к различным типам. Это означает, что жена, в отличие от мужа, Рыцарь. Таким образом, безотносительно к тому, является муж Рыцарем или Плутом, жена должна быть Рыцарем. (Тип «мужа» здесь «недоопределен»; он мог быть Рыцарем, заявляя при этом истинно, что он того же типа, что и его жена, или же он мог быть Плутом, ложно утверждающим, что они с женой одного типа.)

Эта проблема есть специальный случай следующей проблемы: Пусть дано некоторое предложение p, и предположим, что Обитатель острова говорит: «Я есть Рыцарь, если и только если p истинно». Что можно заключить из этого?

Два предложения *эквивалентны*, если они либо оба истинны, либо оба ложны. Другими словами, если одно из них истинно, истинно и другое. Два предложения *неэкви-*

валентны, если они не являются эквивалентными. Итак, Обитатель говорит: «Я есть Рыцарь, если и только если p истинно». Пусть k будет предложением, что Обитатель есть Рыцарь, тогда утверждение Обитателя состоит в том, что k эквивалентно p. Если он – Рыцарь, тогда его утверждение истинно, отсюда k действительно эквивалентно p; и так как k истинно (он – Рыцарь), p также истинно. С другой стороны, если он – Плут, тогда его утверждение ложно; k на самом деле не эквивалентно p, но так как k ложно (он не Рыцарь), p должно быть истинным (потому что предложение, не эквивалентное ложному предложению, должно быть истинным). И поэтому мы видим, что p должно быть истинным, а k недоопределенным. Давайте зафиксируем этот факт в виде Теоремы II.

Теорема II. Пусть дано некоторое предложение p, и предположим, что Обитатель говорит: «Я есть Рыцарь, если и только если p». Тогда p должно быть истинным независимо от того, является Обитатель Рыцарем или Плутом.

Давайте вернемся к проблеме того, имеется ли на острове золото. Предположим, что Обитатель говорит: «Я есть Рыцарь, если и только если на острове есть золото». Тогда согласно Теореме II (где в качестве p выступает предложение, что на острове есть золото) мы видим, что на острове должно быть золото, хотя мы не можем определить, является Обитатель Рыцарем или Плутом.

Мы видим, следовательно, что если Обитатель говорит: «Если я есть Рыцарь, тогда на острове есть золото», согласно Теореме I мы можем вывести как то, что он – Рыцарь, так и то, что на острове есть золото. Но если вместо этого он говорит: «Я есть Рыцарь, если и только если на острове есть золото», тогда согласно Теореме II все, что мы можем вывести, так это то, что имеется золото; мы не можем определить, является ли говорящий Рыцарем или Плутом.

Теорема II есть основание знаменитой проблемы, изобретенной философом Нельсоном Гудменом. Проблема формулируется так: Предположим, что вы прибываете на остров Рыцарей и Плутов и хотите узнать, есть ли на острове золото. Вы встречаете Обитателя, и вам позволено задать только один вопрос, на который можно ответить только «да» или «нет». Какой вопрос вы задали бы ему?

Вопрос, который срабатывает, таков: «Тот ли это случай, что вы есть Рыцарь, если и только если есть золото на острове?» Если ответ «да», тогда, согласно Теореме II на острове есть золото. Если ответ «нет», тогда на острове нет золота (потому что он отрицает, что его рыцарство эквивалентно наличию золота на острове), что равносильно утверждению, что его рыцарство эквивалентно *отсутствию* золота на острове, и поэтому опять-таки согласно Теореме II, на острове нет золота.

5. Некоторые родственные проблемы

Какое утверждение мог бы сделать Обитатель, из которого вы могли бы заключить, что если Обитатель есть Рыцарь, тогда на острове есть золото, но если он – Плут, тогда золото на острове может быть, а может и не быть?

Решение: Есть много утверждений, которые подходят в данном случае. Одно такое утверждение таково: «Я есть Рыцарь, и на острове есть золото». Другое утверждение таково: «На острове есть золото и на острове есть серебро». (Если Обитатель есть Рыцарь, тогда, конечно, есть золото – так же как и серебро, но если есть золото, Обитатель не может быть Рыцарем – не может быть никакого серебра).

6.

Какое утверждение мог бы сделать Обитатель, из которого вы могли бы заключить, что если на острове есть золото, тогда он должен быть Рыцарем, но если на острове нет золота, тогда он мог бы быть либо Плутом, либо Рыцарем?

Решение: Одно из утверждений, которое срабатывает, таково: «Либо я есть Рыцарь, либо на острове есть золото». Фраза «либо – либо» означает *по крайней мере, один – и возможно оба*. И поэтому, если на острове есть золото, тогда определенно истинно, что *либо* Обитатель есть Рыцарь, *либо* на острове есть золото. Следовательно, если на острове есть золото, тогда утверждение Обитателя было истинным, что в свою очередь влечет, что Обитатель должен быть Рыцарем. Это доказывает, что если на острове есть золото, тогда Обитатель есть Рыцарь.

С другой стороны, Обитатель мог бы быть Рыцарем независимо от того, есть ли на острове золото, потому что если он – Рыцарь, тогда истинно, что *либо* он – Рыцарь, *либо* на острове есть золото.

Другое утверждение, которое сработает, таково: «Либо на острове есть золото, либо на острове есть серебро».

7.

Я однажды посетил остров и спросил Обитателя: «На острове есть золото?» Он сказал лишь следующее: «Я никогда не утверждал, что на острове есть золото». Позднее я обнаружил, что на острове все таки *было* золото. Кем был Обитатель – Рыцарем или Плутом?

Решение: Предположим, что Обитатель был Плутом. Тогда его утверждение ложно. Это означает, что когда-то он все-таки делал утверждение, что на острове есть золото. Его утверждение тогда должно быть ложным (так как Плут), что означает, что на острове есть золото. Но я уже сказал, что на острове *было* золото. Отсюда он не может быть Плутом, и значит, он должен быть Рыцарем.

ГЛАВА 4. *В поисках Ооны*

Есть целый архипелаг островов с населением, состоящим из Рыцарей и Плутов, но с одной особенностью – некоторые из его Обитателей являются наполовину-людьми, наполовину-птицами. Эти люди-птицы летают, так же как птицы а говорят как люди.

Существует история о философе – на самом деле, логике, – который посетил этот архипелаг и влюбился в птицу-девушку по имени Оона. Они поженились. Брак был удачным, за исключением того, что жена оказалась любительницей летать! Например, он мог прийти вечером на ужин, но, если это был приятный вечер, она могла улететь на другой остров. Поэтому он должен был мотаться на каноэ от острова к острову для того, чтобы найти Оону и привезти ее домой. Всякий раз, когда Оона приземлялась на остров, Обитатели могли видеть ее в полете и знали, что она приземляется. Но как только она оказывалась на земле, найти ее было трудно. Поэтому первое, что должен был сделать муж по прибытии на остров, это узнать от Обитателей, приземлялась ли Оона. Ситуацию осложняло то, что некоторые из Обитателей были Плутами и не могли утверждать истину. Ниже приведены некоторые инциденты, участником которых был философ.

1.

Однажды муж посетил остров в поисках Ооны и встретил двух Обитателей, А и В. Он спросил их, приземлялась ли на остров Оона. Он получил следующие ответы:

А: Если В и я – оба Рыцари, тогда Оона на острове.
В: Если А и я – оба Рыцари, тогда Оона на острове.

Так находится ли Оона на острове?

Решение: Я ограничу себя более краткими решениями, чем в предыдущих случаях.

Предположим, что А и В – оба Рыцари. Тогда общее утверждение, которое они делают, истинно, и из этого следует, что Оона на острове. Поэтому, если они оба – Рыцари, тогда Оона на острове. Но это как раз то, что они утверждали, так что они оба Рыцари.

2.

В другой раз два Обитателя, А и В, дали следующие ответы:

А: Если любой из нас – Рыцарь, тогда Оона на острове.
В: Это истинно.

Так находится ли Оона на острове?

Решение: Если любой из них является Рыцарем, тогда сделанное им утверждение является истинным, из чего следует, что Оона на острове. Следовательно, если любой их них – Рыцарь, Оона на острове. Таким образом, утверждение, которое они оба сделали, истинно, отсюда, оба они – Рыцари, и поэтому, по крайне мере, один из них есть

Рыцарь. Из этого и истинности утверждения, которое они сделали, следует, что Оона на острове.

3.

Я не помню в точности деталей следующего инцидента. Я знаю только, что логик встретил Обитателей А и В, и что А сказал: «В есть Рыцарь, и Оона на острове». Но я не помню точно, что сказал В. Он сказал либо: «А есть Плут, и Оона не на острове», либо: «А есть Плут, и Оона на острове». Мне так хотелось бы вспомнить это в точности! В любом случае я все же помню, что логик был способен определить, была Оона на острове или нет. Так, была она там или нет?

Решение: Это пример того, что я называю *метапроблемой*. Вам не говорят, что сказал В, но вам говорят, что из того, что сказали А и В, логик смог вывести заключение, была Оона на острове или нет. (Если бы я не сказал вам этого, тогда вы не смогли бы решить проблему!)

Я сначала покажу, что если бы В сказал: «А есть Плут, и Ооны нет на острове», – тогда логик, возможно, не смог бы решить проблему. Поэтому предположим, что В сказал это. Тогда А не мог бы быть, возможно, Рыцарем, потому что если бы он был им, тогда В был бы Рыцарем (как сказал А), что сделало бы А Плутом (как сказал В). Следовательно, А определенно Плут. Но тогда оказалось бы, что либо В – Рыцарь и Ооны нет на острове, либо В есть Плут и Оона на острове, и невозможно сказать, каков из вариантов имеет место. Поэтому, если бы В сказал это, логик не мог бы узнать, есть ли на острове Оона. Но нам дано, что логик все-таки узнал это, и поэтому В *не сказал* этого. Он должен был сказать: «А есть Плут, и Оона на острове». Давайте посмотрим, что получается при этом.

А должен быть Плутом по тем же самым причинам, что и прежде. Если Оона на острове, мы получим следующее противоречие: Тогда истинно, что А есть Плут, и Оона на острове, отсюда, В сделал истинное утверждение, и это делает В Рыцарем. Но тогда А сделал истинное утверждение, что В – Рыцарь и что Оона на острове. Это противоречит тому, что А есть Плут! Единственный выход из противоречия состоит в том, что Ооны *нет* острове. Поэтому Оона не на острове (и, конечно, А и В оба Плуты).

4.

В следующем инциденте логик прибывает на небольшой остров с шестью Обитателями. Он опрашивает каждого, и достаточно любопытно то, что все они говорят одно и то же: «По крайней мере, один Плут на этом острове видел, как Оона вечером приземлялась на острове».

Видел ли какой-либо Обитатель приземление Ооны этим вечером?

Решение: Так как все шесть Обитателей сказали одно и то же, они все являются либо Рыцарями, либо Плутами (все Рыцарями, если сказанное ими истинно, и все Плутами, если сказанное ими ложно). Предположим, что все они – Рыцари. Тогда будет истинным, что, по крайней мере, один Плут на острове видел приземление Ооны, но это невозможно, так как ни один из них не является Плутом. Следовательно, все они должны быть Плутами. Отсюда, сказанное ими является ложным, что означает, что ни один из Плутов не видел приземления Ооны. Но так как все Обитатели Плуты, ни один из них не видел приземления Ооны в этот вечер.

5.

В следующем любопытном инциденте, когда муж прибыл на остров в поисках Ооны, он встретил пять Обитателей – A, B, C, D, E, которые угадали цель его визита и приветствовали его. Они сделали следующие утверждения:

A: Она на острове.
B: Ооны *нет* на острове.
C: Оона была здесь вчера.
D: Ооны нет сегодня здесь, и не было ее и вчера.
E: Либо D есть Плут, или же C есть Рыцарь.

Логик помыслил над этим некоторое время, но так ни к чему и не пришел.

«Не можете ли вы сделать еще какое-нибудь утверждение?» – попросил логик. Тогда A сказал: «Либо E – Плут, либо C – Рыцарь».

Есть ли на острове Оона?

Решение: Мы увидим, что если A есть Плут, получим противоречие: Предположим, что A – Плут. Тогда его второе утверждение ложно, отсюда, E должен быть Рыцарем, а C должен быть Плутом. Так как E – Рыцарь, его утверждение истинно, отсюда, либо D есть Плут, либо C есть Рыцарь. Но C не Рыцарь, поэтому D должен быть Плутом. Так как утверждение D было ложным, либо Оона была здесь вчера, либо она сейчас здесь. Но Оона не была здесь вчера (потому что C сказал это, а C есть Плут), отсюда, Оона на острове сейчас. Но это делает первое утверждение A истинным, что противоречит тому, что A есть Плут! Следовательно, первое утверждение A истинно, и Оона сейчас на острове.

ГЛАВА 5. *Межпланетная путаница*

На Ганимеде – спутнике Юпитера – есть клуб, известный как Марсианско-Венерианский клуб. Его члены либо с Марса, либо с Венеры, хотя иногда допускаются и визитеры. Землянин не может различить венерианцев и марсиан по внешнему виду. Также земляне не могут различить мужчин от женщин с обеих планет по внешнему виду, поскольку они одеты одинаково. Логики, однако, имели некоторое преимущество, так как венерианские женщины всегда говорят правду, а венерианские мужчины лгут. У марсиан все было наоборот: марсиане-мужчины говорят правду, а марсианки лгут.

Однажды Оона и ее муж-логик посетили Ганимед и услышали об этом клубе. «Держу пари, что я сумею отличить женщин от мужчин и марсианцев от венерианцев», – сказал муж гордо своей жене.

– Как ты это сделаешь?

– Мы посетим вечером клуб, и я покажу там, как! – сказал муж (чье имя, кстати говоря, было Джордж).

1.

Они посетили вечером клуб. «Давай посмотрим, что ты можешь, – скептически сказала Оона. – Здесь есть член клуба. Можешь ли ты определить, мужчина это или женщина?» Джордж подошел к члену клуба и задал ему вопрос, на который надо было дать утвердительный или отрицательный ответ. Член клуба ответил, и Джордж тогда определил, мужчина это или женщина, хотя он не мог определить, марсианин это или венерианец.

Какой вопрос был задан?

Решение: Простой вопрос, который срабатывает, таков: «Ты – марсианин?» Предположим, что ответ «да». Отвечающий говорит либо правду, либо ложь. Если это правда, тогда говорящий на самом деле марсианин и, поскольку он утверждает истины, он должен быть мужчиной. Если говорящий лжет, тогда он на самом деле венерианец, отсюда, лгущий – венерианец, отсюда, опять-таки мужчина. Так что в любом случае, ответ «да» указывает на мужской род говорящего. Подобный же анализ (который я оставляю читателю) показывает, что ответ «нет» указывает на то, что говорящий является женщиной.

Конечно, вопрос: «Ты – венерианец?» – работает равно хорошо, и ответ «да» тогда указывает, что говорящий является женщиной, а ответ «нет», что он – мужчина.

2.

«Очень умно!, – сказала Оона после того, как Джордж объяснил решение. – Теперь предположим, что вместо желания узнать, кем является член клуба – женщиной или мужчиной, ты хочешь узнать, с Марса он или с Венеры.

Можешь ли ты сделать это, задав только один вопрос, на который может быть ответ “да” или “нет”?»

«Конечно! – сказал Джордж. – Ты, что, не видишь, как это сделать?»

Оона подумала немного и внезапно нашла решение. Какое именно?

Решение: Вопрос: «Ты – мужчина?» – сработает. (Я оставляю проверку читателю). Альтернативный вопрос: «Ты – женщина?» также сработает.

3.

«Если ты *действительно* умен, – сказала Оона, – ты сможешь определить с помощью только одного вопроса, является ли член клуба мужчиной или женщиной, а также, откуда он. Давай посмотрим, сможешь ли ты сделать это, используя лишь вопрос, на который дается ответ “да” или “нет”!»

«Никто не может *настолько* быть умным!» – ответил Джордж. Почему Джордж сказал это? (Это в существенных чертах повторение последней проблемы из главы 2 моей книги To Mock a Mockingbird.)

Решение: Причина того что невозможно сконструировать вопрос с ответом «да» или «нет» для определения, является ли данный член клуба мужчиной или женщиной и марсианином или венерианцем, состоит в том, что имеются четыре возможности – мужчина-марсианин, женщинамарсианка, венерианский мужчина и венерианская женщина. В то же время, имеются только два возможных ответа на вопрос: «да» или «нет». И поэтому, располагая только двумя возможными ответами, невозможно определить, какая из четырех возможностей имеет место.

4.

И тут проходил член клуба и обронил замечание, из которого Джордж и Оона (которая к тому времени навострилась понимать суть дела) смогли вывести, что он должен быть марсианкой. Какое утверждение сделал этот член клуба?

Решение: Простое утверждение типа: «Я есть венерианский мужчина» – сработает. Ясно, что это утверждение не может быть истинным, потому что мы тогда имели бы противоречие, состоящее в том, что венерианский мужчина делает истинное утверждение. Отсюда, это утверждение ложно, что означает, что говорящий *не* является венерианским мужчиной. Так как утверждение ложно, говорящий должен быть марсианской женщиной.

5.

Следующий проходящий сделал утверждение, из которого Джордж и Оона смогли сделать вывод, что член клуба должен быть венерианкой. Какое утверждение было сделано?

Решение: Эта проблема немного сложнее. Одно из утверждений, которое срабатывает, таково: «Я либо женщина, либо с Венеры». (Заметим, что «либо, либо» означает, по крайней мере, *один или оба*; оно не означает *в точности один.*)

Если это утверждение ложно, тогда говорящий не является женщиной и не с Венеры, отсюда, он должен быть марсианином (мужчиной с Марса). Но мужчина с Марса не делает ложных утверждений, и поэтому мы получаем противоречие. Это доказывает, что утверждение должно быть истинным, отсюда, говорящий (-ая) должен (должна) быть либо женщиной, либо с Венеры, а возможно, тем и другим одновременно. Однако если говорящий (-ая) женщина, она должна быть также с Венеры, а если она с Венеры, она должна быть также женщиной, потому что правду говорящие женщины – с Венеры, а правду говорящие с Венеры – женщины. Следовательно, говорящий (-ая) одновременно с Венеры и женщина.

Кстати говоря, если говорящий делает более сильное утверждение: «Я женщина *и* с Венеры», – невозможно будет определить ни пол, ни расу говорящего (все, что можно вывести при этом, что говорящий не является мужчиной с Марса).

6.

Какое утверждение могло быть сделано мужчиной с Марса, или женщиной с Марса, или мужчиной с Венеры, или женщиной с Венеры?

Решение: Одно из таких утверждений: «Я либо мужчина с Марса, либо женщина с Венеры» – или, проще говоря:

«Я всегда говорю правду». Это мог бы сказать любой лгун или правдолюб.

Вскоре среди членов клуба распространился слух об успехах Джорджа и Ооны в определении расы и пола с помощью логики. Владелец клуба, американский антрепренер по имени Феттер, подошел к столику, где сидели Джордж и Оона, и поздравил их. «Я хотел бы, чтобы вы испытали еще несколько членов клуба, – сказал Феттер, – и посмотреть, как это у вас получается».

7.

Как раз в этот момент мимо проходили два члена клуба. «Присоединяйтесь к нам», – сказал Феттер, который представил подошедших как Орка и Борга. Джордж попросил их рассказать что-нибудь о себе, и они сделали следующие утверждения:

Орк: Борг с Венеры.
Борг: Орк с Марса.
Орк: Борг – мужчина.
Борг: Орк – женщина.

Из этой информации Джордж и Оона смогли успешно идентифицировать пол и расу каждого из них. Кто такие Орк и Борг?

Решение: Предположим, что Орк говорит правду. Тогда Борг будет мужчиной с Венеры, отсюда, Борг должен быть лгуном. Предположим, с другой стороны, что лжет

Орк. Тогда Борг не является мужчиной и не с Венеры. Поэтому Борг должен быть женщиной с Марса, и поэтому Борг должен быть лгуном. Тем самым показано, что независимо от того, сказал Орк правду или нет, Борг определенно лгал.

Так как Борг лгал, Орк не является женщиной и не с Марса, отсюда, Орк должен быть мужчиной с Венеры. Следовательно, Орк также лгал, что означает, что Борг должен быть женщиной с Марса. И поэтому решение таково: Орк есть мужчина с Венеры, а Борг есть женщина с Марса (и все четыре утверждения ложны).

8.

«Вы знаете, – сказал Феттер, после того как Орк и Борг удалились, – что марсиане и венерианцы часто вступают в смешанные браки и у нас в клубе есть несколько таких пар. Вот сейчас к нам подходит пара членов клуба. Давайте просмотрим, можете ли сказать, являются ли они смешанной парой».

Я не помню имен этой пары, так что просто назову их А и В.

– Откуда вы? – спросила Оона члена клуба А.
– С Марса, – был ответ.
– Это не правда! – сказал В.

Так, это смешанная пара или нет?

Решение: Так как А заявил, что он с Марса, А должен быть мужчиной (как мы видели это из решения Проблемы 1), и отсюда, В должен быть женщиной. Если А правдив, тогда А должен быть на самом деле с Марса и В солгал, и, будучи лгуном, должен быть также с Марса. Если А лжет, А на самом деле с Венеры, В говорит правду и, будучи женщиной, также с Венеры. Следовательно, это не смешанная пара, и оба они с одной планеты.

9.

«А вот и другая пара, – сказал Феттер. – И опять я не скажу вам, смешанная это пара или нет. Посмотрим, сможете ли вы угадать, кто из них муж».

Мы назовем их А и В. Джордж спросил: «Вы оба с одной планеты?» Вот их ответы:

А: Мы оба с Венеры.
В: Это не правда!

Который из них муж?

Решение: Если утверждение А истинно, тогда они оба с Венеры, и отсюда, А с Венеры – и А должен быть женщиной. Предположим, что утверждение А ложно. Тогда, по крайней мере, один из них с Марса. Если А с Марса, А должен быть женщиной (так как утверждение А ложно). Если В с Марса, В должен быть мужчиной (так как утверждение В истинно). Отсюда опять-таки А должен быть женщиной. И поэтому А – это жена, а В – муж.

10.

«А вот еще одна пара, – сказал Феттер. – И опять я не скажу вам, смешанная это пара или нет. Посмотрим, что будет!»

В этот раз я вспомнил их имена – это были Ял и Торк.

– Откуда вы? – спросил Джордж каждого.
– Моя половина с Марса, – ответил Торк.
Мы оба с Марс, – сказал Ял.

Это позволило Джорджу и Ооне классифицировать обоих полностью. Который из них муж, и кто жена, и с какой планеты каждый из них?

Решение: Предположим, что Ял сказал правду. Тогда оба на самом деле с Марса, отсюда, утверждение Торка, что Ял с Марса, истинно. Мы, таким образом, имеем невозможность пары с одной планеты, *оба* члена которой говорят правду. Этого не может быть, отсюда, Ял должен быть лгуном. Следовательно, по крайней мере, один из них с Венеры.

Если Ял с Марса, тогда Торк должен быть с Венеры. Но тогда Торк сказал правду, утверждая, что Ял с Марса, и поэтому Торк должен быть женщиной, и отсюда, Ял должен быть мужчиной, и мы имеем невозможность того, что мужчина с Марса утверждает ложь. Следовательно, Ял не может быть с Марса, и должен быть с Венеры. Так как Ял лгал и также с Венеры, он должен быть мужчиной. Также, так как Ял не с Марса, Торк лгал. Отсюда, Торк – лгущая женщина, и значит, с Марса.

Итак, Ял – мужчина с Венеры, а Торк – женщина с Марса.

ЧАСТЬ III

РЫЦАРИ, ПЛУТЫ И ПРОПОЗИЦИОНАЛЬНАЯ ЛОГИКА

ГЛАВА 6. *Элементы пропозициональной логики*

Загадки с правдолюбцами и лгунами из последних трех глав имеют дополнительную значимость, если их рассматривать с точки зрения того, что называется пропозициональной логикой (как мы увидим это в следующей главе). В этой главе мы займемся основами этой логики – логическими связками, истинностными таблицами и тавтологиями. Читатели, уже знакомые с этими понятиями, могут пропустить этот материал и перейти к следующей главе (или же просто освежить свои знания, кратко просмотрев материал).

Логические связки

Пропозициональная логика, подобно алгебре, имеет свой собственный символизм, который легко освоить.

В алгебре буквы x, y, z стоят для неспецифицированных чисел; в пропозициональной логике мы используем буквы p, q, r, s (иногда с индексами) для обозначения неспецифицированных предложений.

Предложения могут соединяться с помощью так называемых *логических связок*. Главными из них являются:

(1) $\neg$ (не)
(2) & (и)
(3) $\vee$ (или)
(4) $\supset$ (если, то)
(5) $\equiv$ (если и только если)

Ниже приведены объяснения логических связок.

(1) Отрицание. Для любого предложения p, под $\neg$p понимается противоположное предложению p, или противоречащее ему. Мы читаем $\neg$p как «не тот случай, что p», или, тот более кратко, «не-p». Предложение $\neg$p называется отрицанием p; оно истинно, если p ложно, и оно ложно, если p истинно. Мы можем суммировать эти два факта в следующей таблице, которая называется *истинностной таблицей* для отрицания. В этой таблице (и во всех таблицах далее) мы используем букву «T» для *истины* и «F» – для *лжи*.

p	$\neg$p
T	F
F	T

Первая строка таблицы говорит, что если p имеет значение T (то есть если p истинно), тогда $\neg$p имеет значение F. Вторая строка говорит, что если p имеет значение F, тогда $\neg$p имеет значение T. Мы можем записать это следующим образом:

$$\neg T = F$$
$$\neg F = T$$

(2) Конъюнкция. Для любых предложений p и q предложение, что p и q оба истинны, пишется в виде «p&q» (иногда «p$\wedge$q»). Мы называем p&q *конъюнкцией* p и q. Она истинна, если оба предложения p и q истинны, но ложна, если одно из предложений ложно. Мы, таким образом, имеем следующие четыре закона конъюнкции:

$$T\&T = T$$
$$T\&F = F$$
$$F\&T = F$$
$$F\&F = F$$

Таким образом, мы имеем истинностную таблицу для конъюнкции:

p	q	p&q
T	T	T
T	F	F
F	T	F
F	F	F

(3) **Дизъюнкция**. Для любых предложений p и q пусть $p \vee q$ будет предложением, что, по крайней мере, одно из предложений p или q истинно. Мы читаем $p \vee q$ как «p или q, либо оба». (Имеется другой смысл «или», а именно: *точно один*, но это не тот смысл, в котором мы используем слово «или». Если p и q оба оказываются истинными, предложение $p \vee q$ считается истинным.) Предложение $p \vee q$ называется дизъюнкцией p и q. Дизъюнкция имеет следующую истинностную таблицу:

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

Мы видим, что $p \vee q$ ложно только в четвертом случае – когда p и q оба ложны.

(4) **Если, то**. Для любых предложений p и q мы пишем $p \supset q$, когда либо p ложно, или p и q оба истинны. Другими словами, если p истинно, тогда истинно и q. Мы иногда читаем $p \supset q$ как «если p, то q», или «p влечет q», или «не тот случай, что p истинно, а q ложно». Мы называем $p \supset q$ импликацией, таблица которой выглядит так:

p	q	$p \supset q$
T	T	T
T	F	F
F	T	T
F	F	T

Мы замечаем, что $p \supset q$ ложно только во второй строке – когда p истинно, а q ложно. Это, вероятно, нуждается в объяснении: $p \supset q$ есть предложение, что не тот случай, когда p истинно, а q ложно. Единственный случай, когда это предложение может быть ложным, это *тот* случай, когда p истинно, а q ложно.

(5) **Если и только если**. Наконец, пусть $p \equiv q$ будет предложением, что p и q или оба истинны, или оба ложны; другими словами, если одно из них истинно, то истинно и другое. Мы читаем $p \equiv q$ как «p истинно, если и только если q истинно», или же как «p и q эквивалентны». (Мы вспомним, что два предложения эквивалентны, если они либо одновременно истинны, либо одновременно ложны.) Вот истинностная таблица для эквивалентности:

p	q	$p \equiv q$
T	T	T
T	F	F
F	T	F
F	F	T

Скобки. Нам нужны скобки для того, чтобы избежать неоднозначности. Например, я пишу $p\&q\vee r$. Читатель не может сказать, имею ли в виду, что p истинно и либо q, либо r истинно, или же я имею в виду, что либо p и q оба истинны, или же r истинно. Если я имею в виду первое, я должен написать $p\&(q\vee r)$; если я имею в виду второе, тогда я должен написать $(p\&q)\vee r$.

Составные истинностные таблицы. Под истинностным значением предложения понимается его истинность или ложность, то есть T, если p истинно, и F, если p ложно. Таким образом, предложения «2 + 2 = 4» и «Лондон есть столица Англии», хотя и являются различными предложениями, имеют одно и то же истинностное значение, а именно, T.

Рассмотрим два предложения – p и q. Если мы знаем истинностное значение p и истинностное значение q, можем определить истинностное значение $p\&q$, $p\vee q$, $p\supset q$, $p\equiv q$, а также истинностное значение $\neg p$ (так же как и истинностное значение $\neg q$). Отсюда следует, что если дана любая комбинация p и q (то есть любое предложение, выразимое в терминах p и q, с использованием логических связок), мы можем определить истинностное значение этой комбинации, как только заданы истинностные значения p и q. Например, предположим, что A есть предложение $(p\equiv(q\&p))\supset(\neg p\supset q)$. Если даны истинностные значения p и q, мы можем последовательно найти значения $q\&p$, $p\equiv(q\&p)$, $\neg p$, $(\neg p\supset q)$, и наконец, $(p\equiv(q\&p))\supset(\neg p\supset q)$. Существуют четыре возможных распределения истинностных значений для p и q, и в каждом из четырех случаев мы можем определить истинностное значение A. Мы можем делать это систематически путем конструирования следующей таблицы:

p	q	q&p	p≡(q&p)	¬p	¬p⊃q	(p≡(q&p))⊃(¬p⊃q)
T	T	T	T	F	T	T
T	F	F	F	F	T	T
F	T	F	T	T	T	T
F	F	F	T	T	F	F

Мы видим, что A истинно в трех случаях и ложно в четвертом.

Давайте рассмотрим другой пример. Пусть B=(p⊃q)⊃ ⊃(¬q⊃¬p), и давайте сконструируем истинностную таблицу для B.

p	q	¬p	¬q	p⊃q	¬q⊃¬p	(p⊃q)⊃(¬q⊃¬p)
T	T	F	F	T	T	T
T	F	F	T	F	F	T
F	T	T	F	T	T	T
F	F	T	T	T	T	T

Мы видим, что B истинно во всех четырех случаях и является примером того, что называется *тавтологией*.

Мы также можем сконструировать истинностную таблицу для комбинации из трех пропозициональных переменных – p, q, r, но в этом случае мы имеем уже восемь случаев для рассмотрения (потому что есть четыре распределения T и F для p и q и для каждого из этих распределений есть две возможности для r). Например, пусть G есть выражение p&(q⊃r))&(r&¬p). Оно будет иметь следующую истинностную таблицу:

p	q	r	q⊃r	p&(q⊃r)	¬p	r&¬p	p&(q⊃r))&(r&¬p)
T	T	T	T	T	F	F	F
T	T	F	F	F	F	F	F
T	F	T	T	T	F	F	F
T	F	F	T	T	F	F	F
F	T	T	T	F	T	T	F
F	T	F	F	F	T	F	F
F	F	T	T	F	T	T	F
F	F	F	T	F	T	F	F

Мы видим, что G ложно во всех восьми случаях; оно противоположно тавтологии и является примером того, что называется противоречием. Не существует таких предложений p, q и r, что p&(q⊃r))&(r&¬p) истинно. (Мы могли бы увидеть это и без таблицы, прибегая просто к здравому смыслу. Предположим, что p&(q⊃r) истинно. Как может тогда (r&¬p) быть истинным, поскольку ¬p ложно?)

Если мы сконструируем истинностную таблицу для четырех переменных (скажем, p, q, r и s), у нас будет 16 случаев для рассмотрения и в таблице будет 16 строк. В общем, для любого целого положительного числа n истинностная таблица для выражения с n переменными должна иметь 2^n строк (каждый раз, когда мы добавляем переменную, число строк удваивается).

Тавтологии

Предложение называется *тавтологией*, если оно установлено на основании только лишь правил истинностных таблиц для логических связок. Например, предположим, что человек говорит, что завтра будет дождь, а второй человек – что не будет. Вряд ли истинностные таблицы помогут нам узнать, кто из них прав. Мы должны дождаться этого завтра и *наблюдать* за погодой. Но предположим, что третий человек говорит сегодня: «Завтра дождь будет или не будет». Вот это я могу *назвать* безопасным предсказанием! Не ожидая наступления этого завтра и не делая никаких наблюдений, я знаю, исходя из *чистого разума*, что он прав. Его утверждение имеет форму $p\&\neg p$ (где p есть утверждение, что завтра будет дождь), и для *каждого* утверждения p предложение $p\&\neg p$ должно быть истинным (как это видно из истинностной таблицы для него).

Более обычное определение тавтологии включает понятие формулы. Под формулой имеется в виду некоторое выражение, построенное из символов $\neg$, $\supset$, $\equiv$, $\vee$ и $\&$, а также пропозициональных переменных p, q, r..., правильно сгруппированных с помощью скобок. Вот точные правила для конструирования формул:

(1) Любая пропозициональная переменная, стоящая отдельно, есть формула.

(2) Для уже сконструированных формул X и Y выражения $(X\vee Y)$, $(X\equiv Y)$, $(X\supset Y)$, $(X\&Y)$, а также $\neg X$ являются формулами.

Следует понимать, что никакое другое выражение не является формулой, если оно не сконструировано согласно правилам (1) и (2) выше.

Когда формула стоит отдельно, мы можем опускать внешние скобки, не рискуя получить неоднозначность – например, когда мы говорим «формула $p\supset q$», мы имеем в виду «формула $(p\supset q)$».

Сама по себе формула не является ни истинной, ни ложной и становится истинной или ложной при интерпретации пропозициональных переменных, которые стоят для определенных предложений. Например, если я спрашиваю: «Является ли формула (p&q) истинной?», вы, вероятно (и будете правы) ответите: «Это зависит от того, какие предложения представляют переменные "p" и "q"». И поэтому такая формула, как «p&q», иногда истинна, а иногда ложна. С другой стороны, такая формула, как «p&¬p», *всегда* истинна (она истинна независимо от того, какое предложение представляет переменная «p», и соответственно называется *тавтологической* формулой. Таким образом, тавтологическая формула есть по определению формула, которая *всегда* истинна. Другими словами, чья истинностная таблица содержит одни лишь T в последнем столбце. Мы можем тогда определить *предложение* как тавтологию, если оно выражается некоторой тавтологической формулой при любой интерпретации пропозициональных переменных. (Например, предложение, что идет дождь или не идет дождь, выражается формулой $p \vee \neg p$, если мы интерпретируем «p» как предложение, что идет дождь.

Логическое следствие и эквивалентность. Пусть даны два предложения X и Y, мы говорим, что X *логически* влечет Y, или что Y есть *логическое следствие* X, если $X \supset Y$ есть тавтология. Мы говорим, что X логически эквивалентно Y, если $X \equiv Y$ есть тавтология; или, что то же самое, если X логически влечет Y и Y логически влечет X.

Некоторые тавтологии

Истинностная таблица есть систематический метод проверки тавтологий, но многие тавтологии могут быть легко распознаны с помощью обычного здравого смысла. Вот несколько примеров:

(1) $((p \supset q) \& (q \supset r)) \supset (p \supset r)$.

Предложение утверждает, что если p влечет q и если q влечет r, тогда p влечет r. Это, конечно, самоочевидно (но может, конечно, быть проверено истинностной таблицей). Эта тавтология имеет специальное название – *силлогизм*.

(2) $(p \& (p \supset q)) \supset q$.

Это означает, что если p истинно и если p влечет q, тогда q истинно. Это иногда перефразируется так: «Все, что влечется истинным предложением, истинно».

(3) $((p \supset q) \& \neg q) \supset \neg p$.

Это означает, что если p влечет ложное предложение, тогда p должно быть ложным.

(4) $((p \supset q) \& (p \supset \neg q)) \supset \neg p$.

Это предложение говорит, что если p влечет q, а также p влечет не-q, тогда p должно быть ложным.

(5) $((\neg p \supset q) \& (\neg p \supset \neg q)) \supset p$.

Этот принцип называется reductio *ad absurdum*. Для того чтобы показать, что p истинно, достаточно показать, что $\neg p$ влечет некоторое предложение как q, так и его отрицание $\neg q$.

(6) $((p \vee q) \& \neg p) \supset q$.

Это знакомый принцип логики: если, по крайней мере, p или q истинно и если p ложно, тогда именно q должно быть истинным.

(7) $((p \vee q) \& ((p \supset r) \& (q \supset r))) \supset r$.

Это другой знакомый принцип, известный под названием доказательства *методом отдельных случаев*. Предположим, что $p \vee q$ истинно. Предположим также, что p влечет q и q влечет r. Тогда r должно быть истинным (независимо от того, является истинным p, или же q, или же оба).

Читатель, не имеющий опыта с пропозициональной логикой, должен выиграть от решения следующих упражнений.

Установить, какое из предложений является тавтологией:

(a) $(p \supset q) \supset (q \supset p)$
(b) $(p \supset q) \supset (\neg p \supset \neg q)$
(c) $(p \supset q) \supset (\neg q \supset \neg p)$
(d) $(p \equiv q) \supset (\neg p \equiv \neg q)$
(e) $\neg (p \supset \neg p)$
(f) $\neg (p \equiv \neg p)$
(g) $\neg (p \& q) \supset (\neg p \& \neg q)$
(h) $\neg (p \vee q) \supset (\neg p \vee \neg q)$
(i) $(\neg p \vee \neg q) \supset \neg (p \vee q)$
(j) $\neg (p \& q) \equiv (\neg p \vee \neg q)$
(k) $\neg (p \vee q) \equiv (\neg p \& \neg q)$
(l) $(q \equiv r) \supset ((p \supset q) \equiv (q \supset r))$
(m) $(p \equiv (p \& q)) \equiv (q \equiv (p \vee q))$

Ответы: (a) Нет, (b) Нет, (c) Да, (d) Да, (e) Нет!, (f) Да, (g) Нет, (h) Да, (i) Нет, (j) Да, (k) Да, (l) Да, (m) Да (оба предложения $p\equiv(p\&q)$ и $(q\equiv(p\vee q))$ эквивалентны $p\supset q$).

Относительно (e) многие начинающие полагают, что никакое предложение p не может влечь свое отрицание. Но это не так! Если p оказывается ложным, тогда $\neg p$ истинно, отсюда, в этом случае $\neg(p\supset\neg p)$ истинно. Однако никакое предложение не может быть *эквивалентным* своему собственному отрицанию, и поэтому (f) есть в самом деле тавтология.

Обсуждение. Значение тавтологий состоит в том, что они не только истинны, но и *логически достоверны*. Никакой научный эксперимент не необходим для установления их истинности – они могут быть проверены только на основании *чистого разума*.

Тавтологии можно альтернативно охарактеризовать без обращения к понятию формулы. Давайте определим *состояние дел* как некоторую классификацию всех предложений на две категории – *истинные* предложения и *ложные* предложения – классификацию, подверженную ограничению, что она должна выполнять истинностные таблицы для логических связок (например, мы не можем классифицировать $p\vee q$ как истинное предложение, если p и q классифицированы как ложные предложения). Тавтология тогда есть предложение, которое истинно в *каждом возможном состоянии дел*.

Это напрямую связано с понятием других возможных миров, принадлежащему Лейбницу. Он утверждал, что из всех возможных миров наш мир является наилучшим. Честно говоря, я понятия не имею, был он прав в этом отношении или нет, но интерес представляет то обстоятельство, что он рассматривал другие возможные миры. Из этого понятия развилась в последние годы целая область философской логики, известной под названием семантики возможных миров. Особый интерес представляет вклад в нее философа Сола Крипке, который мы обсудим в последующих главах. Если задан некоторый возможный мир, множество всех предложений, истинных в этом мире, вместе с множеством предложений, ложных в этом мире, составляют положение дел, имеющее место в *этом* мире. Тавтология тогда истинна не только для этого мира, но для *всех* возможных миров. Физические науки заинтересованы в состоянии дел, которое имеет место в *действительном* мире, в то время как математика и логика изучают *все* возможные состояния дел.

ГЛАВА 7. *Рыцари, Плуты и пропозициональная логика*

Рыцари и Плуты пересмотренные

Мы можем сейчас ввести простое, но фундаментальное устройство, посредством которого огромное число проблем с лгунами и правдолюбами может быть сведено к проблемам в пропозициональной логике. Это устройство будет критически важным в последующих главах.

Давайте вернемся на Остров Рыцарей и Плутов. Пусть есть Обитатель P, пусть k есть утверждение, что P есть Рыцарь. Теперь, предположим, что P делает утверждение X. В общем случае мы не знаем, является ли Обитатель Рыцарем или Плутом, и не знаем также, является ли X истинным или ложным. Но вот что мы знаем определенно: если P есть Рыцарь, тогда X истинно, и обратно, если X истинно, тогда P есть Рыцарь (потому что Рыцари никогда не делают ложных утверждений). И поэтому мы знаем, что P есть Рыцарь, если и только если, X истинно; другими словами, мы знаем, что предложение $k \equiv X$ истинно. И поэтому переводим «P утверждает X» как «$k \equiv X$».

Иногда в ситуацию вовлечены больше двух Обитателей, например мы предполагаем, что имеются два Обитателя – P_1 и P_2. Пусть k_1 будет предложение, что P_1 есть

Рыцарь, и путь k_2 есть предложение, что P_2 есть Рыцарь. Если появляется третий Обитатель, пусть k_3 будет предложением, что P_3 есть Рыцарь. И так далее, пока в рассмотрение не будут включены все Обитатели. Мы тогда переводим: «P_1 утверждает X» как «$k_1 \equiv X$»; «P_2 утверждает X» как «$k_2 \equiv X$», и так далее.

Давайте посмотрим теперь на первую проблему в главе 3. Есть два Обитателя P_1 и P_2 (муж и жена). P_1 утверждает, что P_1 и P_2 оба Плуты. Нам нужно определить, к каким типам относятся P_1 и P_2. Поскольку k_1 есть предложение, что P_1 есть Рыцарь, $\neg k$ эквивалентно предложению, что P_1 есть Плут (так как каждый Обитатель Острова есть либо Рыцарь, либо Плут, но не тот и другой одновременно). Подобным же образом, $\neg k_2$ есть предложение, что P_2 есть Плут. Отсюда, предложение, что P_1 и P_2 являются Плутами, имеет форму $\neg k_1 \& \neg k_2$. P_1 утверждает предложение $\neg k_1 \& \neg k_2$. Теперь, используя наше устройство перевода, *реальность* ситуации такова, что $k_1 \equiv (\neg k_1 \& \neg k_2)$ истинно. Поэтому проблема может быть представлена в чисто пропозициональных терминах: Пусть даны два предложения k_1 и k_2 такие, что $k_1 \equiv (\neg k_1 \& \neg k_2)$ истинно. Каковы при этом истинностные значения k_1 и k_2? Если мы сделаем таблицу истинности, мы увидим, что единственный случай, в котором $k_1 \equiv (\neg k_1 \& \neg k_2)$ истинно, это когда k_1 ложно, а k_2 истинно. (Мы также убедились в этом интуитивно при решении этой проблемы в главе 3). В результате мы имеем $(k_1 \equiv (\neg k_1 \& \neg k_2)) \supset \neg k_1$ в качестве тавтологии, и то же мы имеем и для выражения

$(k_1 \equiv (\neg k_1 \& \neg k_2)) \supset k_2$.

Полное математическое содержание этой проблемы состоит в том, что для любых двух предложении k_1 и k_2 следующее предложение есть тавтология: $(k \equiv (\neg k_1 \& \neg k_2)) \supset \supset (\neg k_1 \& k_2)$.

Читатель мог заметить, что обратное предложение $(\neg k_1 \& k_2) \supset (k_1 \equiv (\neg k_1 \& \neg k_2))$ также является тавтологией; отсюда, предложение $k_1 \equiv (\neg k_1 \& \neg k_2)$ логически *эквивалентно* предложению $(\neg k_1 \& k_2)$.

Давайте посмотрим на вторую проблему в главе 3. Здесь P_1 утверждает, что, по крайней мере, один из P_1 и P_2 является Плутом. Мы заключили, что P_1 есть Рыцарь, а P_2 – Плут. Математическое содержание этого факта состоит в том, что $(k_1 \equiv (\neg k_1 \vee \neg k_2)) \supset (k_1 \& \neg k_2)$ есть тавтология.

Читатель мог заметить, что обратное утверждение также является тавтологией, и отсюда, предложение $(k_1 \equiv \equiv (\neg k_1 \vee \neg k_2)) \equiv (k_1 \& \neg k_2)$ является тавтологией.

Перевод Проблемы 3 представляет определенный теоретический интерес. В самом деле, давайте рассмотрим ее в более общей форме Теоремы 1 из главы 3. Мы имеем Обитателя P, утверждающего определенное предложение q, что если P есть Рыцарь, тогда q истинно (q могло бы быть предложением, что жена P есть Рыцарь, или что на Острове есть золото, или любым другими предложением). Пусть k будет предложением, что P есть Рыцарь. Таким образом, P утверждает предложение $k \supset q$ и реальность ситуации состоит в том, что $k \equiv (k \supset q)$ истинно. Исходя из этого мы должны определить истинностное значение k и q. Как мы видели, k и q оба должны быть истинными. Таким образом, математическое содержание Теоремы 1 из главы 3 состоит в том, что $(k\equiv(k\supset q))\supset(k\&q)$ есть тавтология. Конечно, этот факт на самом деле не зависит от конкретной природы предложения k; для любого предложения, для любых предложений p и q предложение $(p\equiv(p\supset q))\supset(p\supset q)$ есть тавтология. Обратное предложение $(p\&q)\supset(p\equiv(p\supset q))$ также есть тавтология, потому что если $p\&q$ истинно, p и q оба истинны, и отсюда, $p\supset q$ должно быть истинным, отсюда, $p\equiv(p\supset q)$ должно быть истинным. Истинной тогда должна быть и тавтология $p\equiv(p\supset q))\equiv(p\&q)$.

Давайте рассмотрим еще раз Проблему 4 или, лучше, Теорему II из главы 3. Здесь мы имеем Обитателя P, утверждающего, что P есть Рыцарь, *если и только если,* q. Вновь пусть k будет предложением, что P есть Рыцарь, и поэтому P утверждает $k\equiv q$. Следовательно, мы знаем, что $k\equiv(k\equiv q)$ истинно. Их этого мы можем определить, что q должно быть истинным, и поэтому суть математического содержания Теоремы II из главы 3 состоит в том, что $k\equiv(k\equiv q)$ есть тавтология.

Эта тавтология есть то, что я назвал бы *тавтологией Гудмена,* так как она возникает из Проблемы Гудмена, которая обсуждалась в главе 3.

Упражнение 1.

Давайте рассмотрим трех Обитателей P_1, P_2 и P_3. Предположим, что P_1 и P_2 делают следующие утверждения:

P_1: P_2 и P_3 оба Рыцари.

P_2: P_1 есть Плут, и P_3 есть Рыцарь.

К каким типам принадлежат P_1, P_2 и P_3?

Упражнение 2.

(a) Является ли тавтологией следующее предложение:

$$((k\equiv(k_2\&k_3))\&(k_2\equiv(\neg k_1\&k_3)))\supset(\neg k_1\&\neg k_2)\&\neg k_3?$$

(b) Как это упражнение соотносится с Упражнением 1?

Упражнение 3.

Покажите, что предложение p_3 есть логическое следствие двух следующих предложений:

(1) $p_1 \equiv \neg p_2$;

(2) $p_2 \equiv (p_1 \equiv \neg p_3)$.

Упражнение 4.

Предположим, что P_1, P_2, P_3 являются тремя Обитателями Острова Рыцарей и Плутов и что P_1 и P_2 делают следующие утверждения:

P_1: P_2 есть Плут;

P_2: P_1 и P_3 принадлежат к различным типам.

(a) Является ли P_3 Рыцарем или Плутом?

(b) Как это соотносится с Упражнением 3?

Ответы на Упражнения 1–4.

1. Все трое являются Плутами.
2. Предложение является тавтологией.
3. Мы оставляем это читателю.
4. P_3 есть рыцарь.

Проблемы Ооны

Многие из проблем Ооны из главы 4 могут быть также решены с помощью таблиц истинности. Рассмотрим, например, первую проблему: мы имеем двух Обитателей P_1 и P_2; P_1 утверждает, что если P_1 и P_2 оба Рыцари, тогда Оона на острове. P_2 утверждает то же самое. Пусть O будет предложением, что Оона на острове. Используя устройство перевода, мы знаем, что два следующих предложения истинны:

$$(1)\ k_1 \equiv ((k_1 \& k_2) \supset O);$$
$$(2)\ k_2 \equiv ((k_1 \& k_2) \supset O).$$

Мы должны определить, истинно или ложно предложение O. Ну, если вы выпишите истинностную таблицу для конъюнкции (1) и (2), то есть для предложения

$$(k_1 \equiv ((k_1 \& k_2) \supset O)) \& (k_2 \equiv ((k_1 \& k_2) \supset O)),$$

вы увидите, что последний столбец содержит T в тех и только тех местах, в которых O имеет значение T. Следовательно, O должно быть истинным.

Упражнение 5.

Предположим, что муж ищет Оону и встречает двух Обитателей A и B, которые делают следующие утверждения:

A: Если B есть Рыцарь, тогда Оона не находится на острове.

B: Если A есть Плут, тогда Оона не находится на острове.

Находится ли Оона на острове?

Упражнение 6.

На другом острове два Обитателя A и B делают два следующих заявления:

A: Если хотя бы один из нас Рыцарь, тогда Оона на острове.

B: Если хотя бы один из нас Плут, тогда Оона на острове.

Находится ли Оона на острове?

Упражнение 7.

На другом острове два Обитателя A и B делают два следующих заявления:

A: Если я – Рыцарь и B – Плут, тогда Оона на острове.

B: Это неверно!

Находится ли Оона на острове?

Ответы на Упражнения 5–7.

5. Оона не на острове (и оба Обитателя являются Рыцарями).
6. Оона на острове (и оба Обитателя являются Рыцарями).
7. Оона на острове (и A – рыцарь, а B – Плут).

Вновь марсиане и венерианцы

Многие из загадок о мужчинах и женщинах с Марса и Венеры из главы 5 могут быть решены с помощью таблиц истинности, хотя тут устройство перевода немного более сложное, и этот тип проблем не рассматривается далее в этой книге.

Давайте перечислим членов клуба в некотором порядке P_1, P_2, P_3 и т.д., и для каждого члена i пусть V_i будет предложением, что P_i является венерианцем, и пусть F_i будет предложением, что P_i является женщиной. Тогда «P_i является марсианином» может быть записано как $\neg V_i$, а «P_i является мужчиной» может быть записано как $\neg F_i$. Теперь P_i говорит истину, если и только если, P_i есть либо венерианская женщина, или же марсианский мужчина, что может быть символизировано как $(V_i \& F_i) \vee (\neg V_i \& \neg F_i)$ или, проще, $V_i \equiv F_i$. И поэтому, если P_i утверждает предложение X, реальность ситуации отражается следующим предложением: $(V_i \equiv F_i) \equiv X$.

Итак, в действии наше устройство перевода. Всякий раз, когда P_i утверждает X, мы пишем $(V_i \equiv F_i) \equiv X$.

Рассмотрим, например, Проблему 7 из главы 5 – случай Орка и Борга. Пусть P_1 будет Орком, а P_2 – Боргом. Тогда мы имеем следующие четыре предложения (после применения устройства перевода):

(1) $(V_1 \equiv F_1) \equiv V_2$;
(2) $(V_2 \equiv F_2) \equiv \neg V_1$;
(3) $(V_1 \equiv F_1) \equiv \neg F_2$;
(4) $(V_2 \equiv F_2) \equiv F_1$.

Истинностные значения V_1, F_1, V_2, F_2 могут быть найдены по таблице. (Есть четыре неизвестных V_1, F_1, V_2, F_2, и поэтому есть 16 случаев для рассмотрения!)

Замечание: Не все проблемы, связанные с утверждениями истинности и ложности, могут быть решены с помощью устройства перевода. Это устройство работает превосходно для решения проблем, в которых нечто утверждается говорящим, и мы должны тогда дедуцировать определенные факты о них. Но для более трудных типов задач, в которых мы должны изобрести вопрос или утверждение для определенных целей, нужны большие усилия. Есть другие систематические устройства, которые помогают во многих случаях, но этот вопрос выходит за пределы главной темы этой книги.

ГЛАВА 8. *Логическое замыкание и непротиворечивость*

Определение одних логических связок через другие логические связки

Мы сейчас знакомы с пятью основными логическими связками: $\neg$, &, $\vee$, $\supset$, $\equiv$. Мы могли бы начать обходиться меньшим числом связок и определить одни логические связки в терминах других. Это может быть сделано несколькими способами, некоторые из них будут иллюстрироваться следующими загадками.

1.

Предположим, что разумное существо с Марса прибывает на Землю и хочет научиться нашей логике. Он заявляет, что понял слова «нет» и «и», но не знает значения слова «или». Как можно его объяснить ему, используя только понятия *нет* и *и*?

Давайте перефразируем проблему. Пусть дано некоторое предложение p, разумное существо с Марса знает значение $\neg$p, и если даны некоторые предложения p и q, он знает значения p&q. Марсианин ищет способ написать выражение или формулу с буквами p и q, используя только логические связки $\neg$ и & такие, что выражение эквивалентно выражению $p \vee q$. Как это можно сделать?

Решение: Сказать, что, по крайней мере, одно из двух предложений p и q истинно, значит, сказать, что не тот случай, что p и q оба ложны; другими словами, не тот случай, что $\neg p$ и $\neg q$ оба истинны. Так что $p \vee q$ эквивалентно предложению $\neg(\neg p \& \neg q)$. Так как Марсианин понимает $\neg$ и &, он поймет $\neg(\neg p \& \neg q)$. И поэтому вы можете сказать Марсианину: «Когда я говорю p или q, все, что я имею в виду, что не тот случай, что не-p и не-q».

2.

Теперь, предположим, что к нам прибывает леди с планеты Венера и говорит нам, что она понимает значение $\neg$ и $\vee$, но хочет, чтобы объяснили ей значение &. Как определить & в терминах $\neg$ и $\vee$? (То есть какое выражение в терминах p, q, $\neg$ и $\vee$ логически эквивалентно $p \& q$?)

Решение: Сказать, что p и q оба истинны, эквивалентно тому, что не тот случай, что либо p, либо q ложны, другими словами, не тот случай, что либо $\neg p$, либо $\neg q$ истинны. И поэтому, $p \& q$ логически эквивалентно $\neg(\neg p \vee \neg q)$.

3.

В этот раз прибыло существо с Юпитера, которое понимает $\neg$, &, $\vee$ и хочет от нас, чтобы мы определили $\supset$ в этих терминах. Как это можно сделать?

Решение: Это можно сделать несколькими способами. С одной стороны, $p \supset q$ логически эквивалентно $\neg p \vee (p \& q)$. Оно также эквивалентно $\neg(p \& \neg q)$ (не тот случай, что p истинно и q ложно) и предложению $\neg p \vee q$.

4.

В этот раз прибыло существо с Сатурна, которое понимает, что довольно странно, $\neg$ и $\supset$, но не понимает ни &, ни $\vee$. Как мы можем объяснить ему & и $\vee$?

Решение: $p \vee q$ логически эквивалентно $\neg p \supset q$. И поэтому мы можем получить $\vee$ в терминах $\neg$ и $\supset$. Раз мы имеем $\neg$ и $\vee$, можем получить & с помощью Решения Проблемы 2 выше. Более прямо, $p \& q$ эквивалентно $\neg(p \supset \neg q)$, в чем читатель может убедиться сам.

5.

Теперь прибыло существо с Урана, которое понимает только связку $\supset$. В этом случае невозможно объяснить ему, что означает &, или же, что означает $\neg$, но возможно определить $\vee$ в терминах точно одной связки $\supset$. Как это сделать? (Решение не очевидно. Как это возможно сделать, стало частью фольклора математической логики, но я не смог найти логика, который открыл это решение).

Решение: Это в самом деле тонкий ход! Предложение $p \vee q$ логически эквивалентно $(p \supset q) \supset q$, что можно проверить по истинностным таблицам.

6.

Ясно, что логическая связка $\equiv$ может быть определена в терминах $\neg$, &, $\vee$. Покажите два различных способа сделать это.

Решение: $p \equiv q$ эквивалентно $(p \supset q) \& (q \supset p)$. Это также эквивалентно $(p \& q) \vee (\neg p \& \neg q)$.

7. Специальная проблема

Предположим, что существо из космоса понимает значение $\supset$ и $\equiv$. Тогда будет невозможно объяснить ему значение $\neg$, но можно объяснить значение &. Как это можно сделать? (Это, насколько я знаю, мое открытие.)

Решение: Мы уже решили эту проблему в предшествующей главе в связи с третьей проблемой Рыцарей и Плутов; мы показали, что $p \& q$ логически эквивалентно $p \equiv (p \supset q)$.

Две другие связки: Мы видели, что логические связки $\equiv$, $\supset$, &, $\neg$, $\vee$ могут быть определены с помощью двух связок $\neg$ и & или же $\neg$ и $\vee$. А есть ли какая-нибудь точно *одна* связка, с помощью которой могут быть определены все остальные связки? Проблема была решена Г. Шеффером в 1913 г. Он определил «$p \mid q$» как означающее, что p и q не являются оба истинными. Этот символ известнее под названием «строка Шеффера» и интерпретируется как «p и q несовместимы» (по крайней мере, одно из них ложно). Он показал, что $\equiv$, $\supset$, &, $\neg$, $\vee$ могут быть определены с помощью одной лишь строки Шеффера. Другая логическая связка, которая дает все остальные связки, есть $\downarrow$. Этот символ используется для обозначения *взаимного отрицания*. Мы читаем $p \downarrow q$ как «p и q оба ложны» или же «ни p, ни q не являются истинными».

8.

Как можно определить $\equiv$, $\supset$, &, $\neg$, $\vee$ с помощью строки Шеффера? Как эти связки могут быть определены с помощью взаимного отрицания?

Решение: Если дана строка Шеффера, мы можем получить остальные связки следующим образом. Прежде всего, $\neg p$ логически эквивалентно $p \mid p$ (Предложение $p \mid p$ говорит, что, по крайней мере, одно из предложений p или p ложно, но так как два предложения p и p есть одно и то же, оно просто говорит, что p ложно). Теперь, раз мы имеем $\neg$, мы можем определить $p \& q$ как $\neg(p \mid p)$. (Так как $p \mid p$ есть предложение, что, по крайней мере, одно из p или q ложно, его отрицание $\neg(p \mid p)$ эквивалентно предложению, что p и q оба истинны). Раз мы имеем $\neg$ и &, можем получить $\vee$ (по Проблеме 1), затем $\supset$ и $\equiv$ (Проблемы 3 и 6). Мы, таким образом, получаем $\equiv$, $\supset$, &, $\neg$, $\vee$ из строки Шеффера.

Если мы начинаем с взаимного отрицания $\downarrow$ взамен строки Шеффера, мы действуем так: сперва возьмем $\neg$ для $p \downarrow p$, затем мы берем $p \vee q$ в качестве $\neg(p \downarrow q)$. Как только мы имеем $\neg$ и $\supset$, мы можем получить все остальное способом, который уже изложен выше.

Логическая константа $\perp$. Предложение X называется логически противоречивым или логически ложным, если его отрицание $\neg X$ является тавтологией. Например, для любого предложения p предложение $(p\&\neg p)$ логически ложно. Также ложно $p \equiv \neg p$.

Мы будем использовать стандартный символ «$\perp$» в качестве представления любой конкретной логической ложности (какой именно, неважно). Этот символ используется далее по всей книге, как и в любой другой книге. (Символ произносится как «ит»; это перевернутый символ T). Для любого предложения p предложение $\perp \supset p$ есть тавтология (потому что $\perp$ логически ложно, и поэтому $\perp \supset p$ истинно независимо от того, истинно p или ложно). Таким образом, *любое* предложение есть логическое следствие $\perp$. (Это хорошо, что $\perp$ не является истинным, потому что если бы оно было истинным, все было бы истинным и весь мир попросту взорвался бы!)

Многие современные формулировки пропозициональной логики строятся просто на основании только лишь двух связок $\perp$ и $\supset$, потому что все остальные связки могут быть получены из этих двух (см. Проблему 9 ниже). Мы примем эту стратегию, потому что она лучше всего согласуется с материалом этой книги. Мы определяем T как $\perp \supset \perp$. Мы будем говорить о $\perp$ как логической ложности и о T как логической истинности (ясно, что T есть тавтология).

9.

Как определить все логические связки через $\perp$ и $\supset$?

Решение: Предложение $\neg p$ логически эквивалентно $p \supset \perp$, и поэтому мы можем получить $\neg$ из $\supset$ и $\perp$. Раз мы имеем $\neg$ и $\supset$, мы можем получить $\vee$ и & (по Проблеме 4). Затем мы можем получить $\equiv$ из $\supset$ и &.

Логическое замыкание

Логически познающая машина. Для иллюстрации важного понятия *логического замыкания* давайте вообразим вычислительную машину, запрограммированную для доказательства различных суждений. Всякий раз когда машина доказывает суждение, она печатает его (более точно, она печатает предложение, которое выражает это суждение). Машина, предоставленная сама себе, будет работать бесконечно.

Мы назовем машину *логически познающей*, если она удовлетворяет следующим двум условиям:

(1) Каждая тавтология рано или поздно будет доказана машиной.

(2) Для двух любых предложений p и q, если машина доказывает p и доказывает $p \supset q$, она рано или поздно докажет q. (Мы могли бы считать машину выводящей q из двух

предложений p и $p \supset q$. Конечно, вывод является общезначимым.)

Логически познающая машина имеет одно очень важное свойство, которое иллюстрируется несколькими примерами.

10.

Предположим, что машина является логически познающей.

(a) Если машина доказывает p, докажет ли она необходимо $\neg\neg p$?

(b) Если машина доказывает p и доказывает q, докажет ли она предложение $p\&q$?

Решение:

(a) Предположим, машина доказывает p. Она также докажет $p \supset \neg\neg p$ (так как это тавтология), отсюда, она напечатает $\neg\neg p$ (по второму условию, определяющему логическое познавание).

(b) Предположим, что машина доказывает p и доказывает q. Далее, предложение $p \supset (q \supset (p\&q))$ является тавтологией, отсюда, машина докажет его. Раз машина докажет p и $p \supset (q \supset (p\&q))$, она должна доказать $q \supset (p\&q)$. Поскольку она доказывает q, она должна доказать $p\&q$.

(На самом деле, эта проблема есть просто специальный случай следующей проблемы, как читатель может убедиться сам.)

Логическое замыкание. В логике есть старое правило, называемое *modus ponens*, которое состоит в том, что если доказано p и доказано $p \supset q$, тогда можно вывести q. Множество C предложений логически замкнуто относительно *modus ponens*, если для любых предложений p и q, если p и $p \supset q$ принадлежат множеству C, тогда q также принадлежит этому множеству.

Мы можем сейчас определить множество C предложений, которое логически замкнуто, если выполняются два условия:

Условие 1. C содержит все тавтологии.

Условие 2. Для любых двух предложений p и q, если p и $p \supset q$ содержатся в C, в C содержится и q.

Таким образом, логически замкнутое множество есть множество, которое содержит все тавтологии и замкнуто относительно *modus ponens*. (Причина названия *логически замкнутое* будет ясна вскоре.)

Сказать, что логическая машина является логически познающей, равносильно тому, что сказать, что множество C всех предложений, которые машина может доказать, логически замкнуто. Однако логически замкнутые множества также возникают в ситуациях, в которых нет никакого участия машины. Например, мы будем рассматривать математические системы, в которых множество всех предложений, доказуемых в системе, является логически замкнутым. Также значительная часть этой книги будет иметь дело с логиками, чье множество вер также логически замкнуто.

Логическое следствие. Мы определяем Y как логическое следствие X, если предложение $X \supset Y$ есть тавтология. Мы будем говорить, что Y есть логическое следствие двух предложений X_1 и X_2, если оно является следствием предложения $X_1 \& X_2$, другими словами, если $(X_1 \& X_2) \supset Y$ есть тавтология, или, что то же самое, если $X_1 \supset (X_2 \supset Y)$ есть тавтология. Мы определим Y как логическое следствие X_1, X_2, X_3, если $((X_1 \& X_2) \& X_3) \supset Y$ есть тавтология, или, что то же самое, если $X_1 \supset (X_2 \supset (X_3 \supset Y))$ есть тавтология. Более обще, для любого конечного множества предложений $X_1, \ldots, X_n$ мы можем определить Y как логическое следствие этого множества, если предложение $(X_1 \& \ldots \& X_n) \supset Y$ есть тавтология.

Важность логически замкнутых множеств состоит в том, что они обладают следующим свойством:

Принцип L (Принцип логического замыкания). Если C логически замкнуто, тогда для любых n предложений X_1, X_2, … & X_n в C все логические следствия этих n предложений также содержатся в C.

Обсуждение. Возвращаясь к нашему примеру логически познающей машины, отметим, что Принцип L говорит нам, что если машина когда-либо докажет предложение p, тогда она рано или поздно докажет все логические следствия p; если машина когда-либо докажет два предложения p и q, она рано или поздно докажет все логические следствия p и q и т.д., для любого конечного множества предложений.

То же самое применимо к логику, чье множество вер логически замкнуто. Если логик когда-либо поверит в p, он рано или поздно поверит во все логические следствия p; если он когда-либо поверит в p и q, он рано или поздно поверит во все логические следствия p и q и т.д.

11.

Почему Принцип L верен?

Решение: Давайте рассмотрим случай $n = 1$. Предположим, что X_1 содержится в C, а Y – логическое следствие X_1. Тогда $X_1 \supset Y$ есть тавтология, отсюда, она содержится в C (по Условию 1). Так как X_1 и $X_1 \supset Y$ оба в C, в C содержится и Y (по Условию 2).

Теперь рассмотрим случай $n = 2$. Предположим, X_1 и X_2 оба содержатся в C, и Y – логическое следствие X_1 и X_2. Тогда $(X_1 \& X_2) \supset Y$ есть тавтология, отсюда, $X_1 \supset (X_2 \supset Y)$ есть тавтология (как может убедиться читатель), и следовательно, содержится в C. Так как X_1 и $X_1 \supset (X_2 \supset Y)$ оба содержатся в C, в C содержится и $X_2 \supset Y$ (Условие 2). Так как X_2 и $X_2 \supset Y$ содержатся в C, в C содержится и Y (опять-таки по Условию 2).

Для случая $n = 3$ предположим, что Y есть логическое следствие X_1, X_2 и X_3. Тогда $X_1 \supset (X_2 \supset (X_3 \supset Y))$ есть тавтология – оно логически эквивалентно $(X_1 \& X_2 \& X_3) \supset Y$. Тогда, используя Условие 2 трижды, мы последовательно получаем $X_2 \supset (X_3 \supset Y)$ в C, $(X_3 \supset Y)$ в C, и наконец, Y в C.

Должно быть ясно, как это доказательство обобщается для любого положительного целого числа n.

Замечание: Мы отмечали, что Проблема 10 есть просто специальный случай настоящей проблемы. Причина этого, конечно, состоит в том, что $\neg\neg p$ есть логическое следствие p, отсюда, любое логически замкнутое множество, содержащее p, должно также содержать $\neg\neg$ p. Также p&q есть логическое следствие двух предложений p и q, поэтому любое логически замкнутое множество, содержащее p и q, должно также содержать p&q.

12.

Другое важное свойство логически замкнутых множеств состоит в том, что если C логически замкнуто и содержит некоторое предложение p и его отрицание $\neg p$, тогда в C должно быть любое предложение.

Почему это так?

Решение: Предположим, что p и его отрицание содержатся в C и что C логически замкнуто. Пусть q будет любое произвольное предложение. Предложение $(p\&\neg p)\supset q$ есть тавтология (это читатель может проверить по истинностной таблице или же, проще, видя, что поскольку $p\&\neg p$ должно быть ложным, $(p\&\neg p)\supset q$ должно быть истинным). И поэтому q есть логическое следствие истинных предложений p и $\neg p$. Тогда в соответствии с Принципом L предложение q также должно быть в C. И поэтому *любое* предложение q содержится в C.

Непротиворечивость

Последняя проблема подводит нас к важному понятию *непротиворечивости.* Логически замкнутое множество C называется *противоречивым*, если оно содержит $\perp$, и *непротиворечивым*, если оно не содержит $\perp$.

Ниже приведено другое важное свойство логически замкнутых множеств.

Принцип C. Предположим, C логически замкнуто. Тогда следующие три свойства эквивалентны друг другу (любое из них влечет два других):

(1) C противоречиво (C содержит $\perp$).

(2) C содержит *все* предложения.

(3) C содержит некоторое предложение p и его отрицание $\neg$p.

Замечание: Пусть дано множество S, не замкнутое логически; S называется противоречивым, если $\perp$ есть логическое следствие некоторого конечного подмножества $X_1, \ldots, X_n$ предложений в S. (Это, между прочим, эквивалентно утверждению, что *любое* предложение есть логическое следствие некоторого конечного подмножества S). Однако мы будем иметь дело почти исключительно с множествами, которые логически замкнуты.

13.

Доказать Принцип C.

Решение: Мы покажем, что три условия эквивалентны, показав, что (1) влечет (2), которое, в свою очередь, влечет (3), которое, в свою очередь, влечет (1).

Предположим (1). Так как любое предложение есть логическое следствие $\perp$ и $\perp$ содержится в C, любое предложение содержится в C (по Принципу L). Таким образом, (2) справедливо.

Совершенно ясно, что (2) влечет (3), потому что если *все* предложения содержатся в C, тогда для любого предложения p и его отрицание $\neg$p содержатся в C.

Сейчас предположим, что (3) справедливо, то есть для некоторого p, p и $\neg$p содержатся в C. Так как $\perp$ есть логическое следование p и $\neg$p, $\perp$ должно быть в C (согласно Принципу L), то есть C должно быть противоречивым.

ЧАСТЬ IV

БУДЕМ ВНИМАТЕЛЬНЫ!

ГЛАВА 9. *Парадоксально?*

Сейчас мы имеем все, чтобы начать наше путешествие к сложным материям с геделевой непротиворечивостью, которые мы постигнем в главе 12, встретив на пути многие интересные проблемы. Давайте начнем с проблемы, тесно связанной с одним из вариантов Парадокса сюрприза в главе 2.

Мы опять на острове Рыцарей и Плутов, где справедливы следующие три утверждения: (1) Рыцари делают только истинные утверждения; (2) Плуты делают только ложные утверждения; (3) каждый Обитатель является либо Рыцарем, либо Плутом. Эти три утверждения могут быть названы «правилами Острова».

Вспомним, что никакой Обитатель не может заявить, что он не Рыцарь, так как Рыцарь никогда не сделает ложного утверждения, что он не Рыцарь, а Плут не сделает истинного утверждения, что он не Рыцарь.

Предположим, что логик наносит визит на Остров и встречает Обитателя, который делает следующее утверждение ему: «*Ты никогда не узнаешь, что я – Рыцарь*».

Получаем ли мы при этом парадокс? Давайте посмотрим. Логик начинает мыслить так: «Предположим, что Обитатель есть Плут. Тогда его утверждение ложно, что означает, что однажды я *узнаю*, что он – Рыцарь, но я не могу знать, что он – Рыцарь, если он на самом деле не окажется Рыцарем. Таким образом, если он – Плут, то отсюда следует, что он – Рыцарь, а это противоречие. Следовательно, он не может быть Плутом, он должен быть Рыцарем».

Пока все хорошо, и мы не имеем парадокса. Но затем логик продолжает мыслить так: «Сейчас я знаю, что он – Рыцарь, хотя он и сказал, что я никогда не узнаю этого. Отсюда, его утверждение ложно, и это означает, что он – Плут. Парадокс!»

Вопрос: Это действительно настоящий парадокс?

Обсуждение: Эта проблема требует весьма добротного анализа! Начать с того, что этот парадокс (если он вообще является таковым) является тем, что, скорее, называется прагматическим парадоксом, а не логическим. Дело в том, что тут включены не только логические понятия типа истинности и ложности, но и прагматические, вроде *знания*. При осознании прагматического характера вопроса нельзя исключать возможность того, что реальная или кажущаяся парадоксальность зависит от человека, к которому обращено утверждение. И это действительно так! В качестве крайнего примера Обитатель мог говорить с мертвым, и тогда никакого парадокса не возникает. (Он указывает на труп и говорит: «Ты никогда не узнаешь, что я – Рыцарь». Ну, он определенно прав; труп никогда не узнает, что тот – Рыцарь, и поэтому Обитатель на самом деле – Рыцарь, так как то, что он сказал, истинно. Но поскольку труп никогда не узнает этого, противоречия не возникает.) Если взять менее крайний пример, Обитатель мог говорить с живым, но глухим человеком, который просто не мог слышать произносимого; и вновь парадокс при этом не возникает.

Поэтому мы должны предположить, что посетитель Острова жив и слышит произносимое утверждение, но этого все еще недостаточно. Мы должны предположить определенные способности посетителя к *размышлению*, потому что если посетитель не имеет такой способности, он не разберется в аргументе, который я привел. (Обитатель мог бы сказать: «Ты никогда не узнаешь, что я – Рыцарь». Посетитель мог бы тогда сказать: «Это очень интересно», – и пойти прочь, не размышляя над тем, что сказано. При этом никакого парадокса не возникало бы.) И поэтому мы должны точно очертить, какого рода способности к размышлению есть у логика.

Мы определим индивида как *Мыслителя типа 1*, если он полностью понимает пропозициональную логику, то есть выполняются два условия:

(1) Он верит во все тавтологии.

(2) Для любых предложений X и Y, если он верит в X и верит в $X \supset Y$, тогда он верит в Y.

В терминологии главы 8 множество вер Мыслителя типа 1 логически замкнуто. Конечно, это предположение в высшей степени идеализировано, поскольку имеется *бесконечно* много тавтологий, отсюда наше предположение влечет что-то вроде бессмертия Мыслителя. Однако такие мелочи не занимают нас в бесконечной сфере математики. Мы просто вообразим себе Мыслителя запрограммированным так, что: (1) рано или поздно он поверит в каждую тавтологию; (2) если он когда-либо поверит в p и даже в $p \supset q$, тогда рано или поздно он поверит в q. Тогда в соответствии с Принципом L из главы 8 имеем: пусть дано конечное множество S предложений, если Мыслитель верит во все предложения в S, тогда для любого предложения Y, которое является логическим следствием S, Мыслитель рано или поздно поверит в Y.

Мы, во-первых, должны предположить определенное *самосознание* у Мыслителя. В частности, мы должны предположить, что если Мыслитель знает вообще что-либо, тогда он знает, что он знает это (в противном случае он никогда не мог бы сказать: «Сейчас я знаю, что он – Рыцарь», – и мой аргумент не мог бы пройти).

Рассматривая проблему при всех этих предположениях, зададим вопрос: возникает ли истинный парадокс? Все еще нет, потому что, хотя я сказал вам, что правила Острова действуют (каждый Обитатель – либо Рыцарь, либо Плут; Рыцари делают только истинные утверждения, Плуты делают только ложные утверждения), мы должны сделать дополнительное предположение, что Мыслитель *верит*, что правила Острова действуют. В самом деле, вполне разумно, что Мыслитель мог бы верить в это с самого начала, но после обнаружения у себя противоречивости в ходе следования аргументу, который я привел, он должен был бы иметь рациональные основания для *сомнения* в правилах Острова. (Я полагаю, что как вы, так и я сделали бы то же самое после попадания в такое затруднительное положение!) Ну, для того чтобы сделать проблему более интересной, давайте сделаем наше последнее предположение, что Мыслитель *верит и продолжает верить* в правила Острова.

1.

Начинаем сейчас исследовать интересную проблему. При дополнительных условиях, которые мы сделали, аргументы Мыслителя, что Обитатель есть Рыцарь и что Обитатель есть Плут, кажутся полностью обоснованными. И все же Обитатель не может быть одновременно Рыцарем и Плутом. Что же не так с аргументацией Мыслителя?

Решение: Я сформулировал проблему выше вводящим в заблуждение образом. (Временами я бываю коварным!) Неправильным был не аргумент Мыслителя; описанная ситуация не могла бы никогда возникнуть. Если Мыслитель типа 1 приходит на Остров Рыцарей–Плутов и верит в правила Острова (и слышит утверждения его Обитателей), тогда логически невозможно, что Обитатель скажет ему: «Ты никогда не узнаешь, что я – Рыцарь».

Для доказательства этого нам даже не нужно одно из уже сделанных предположений, а именно: если Мыслитель знает нечто, тогда он знает, что он знает нечто. Даже без этого предположения мы можем получить противоречие следующим образом. Мыслитель думает так: «Предположим, Обитатель есть Плут. Тогда его утверждение ложно, и значит, я узнаю, что он – Рыцарь, что влечет, что он на самом деле Рыцарь. Следовательно, предположение о том, что он Плут, ведет к противоречию, и значит, он должен быть Рыцарем».

Без дальнейшего углубления в процесс размышления логика мы можем вывести противоречие. До сих пор логик размышлял правильно и пришел к заключению, что Обитатель есть Рыцарь. Поскольку он мыслит правильно, Обитатель на самом деле Рыцарь, и поэтому Мыслитель знает, что Обитатель есть Рыцарь. Однако Обитатель говорит, что тот никогда не узнает этого, и отсюда, Обитатель должен быть Плутом. Следовательно, Обитатель одновременно есть Рыцарь и Плут, что и есть противоречие.

Вместо слова «знает» мы могли бы использовать также и слово «правильно верит», и наш аргумент был бы значим. Мы говорим, что индивид *правильно верит* в предложение p, если он верит в p, и p истинно.

Итак, мы только что доказали Теорему 1.

Теорема 1. Пусть имеется остров Рыцарей и Плутов, а также Мыслитель типа 1, который верит в правила Острова (который слышит адресованное ему утверждение). Тогда логически невозможно, чтобы Обитатель сказал Мыслителю: «Ты никогда не поверишь правильно, что я – Рыцарь».

Обсуждение. Некоторые критически настроенные читатели могут оспорить доказательство, которое я дал для этой теоремы, на том основании, что я снабдил Мыслителя большими способностями, чем я приписал ему в явном виде, а именно, что Мыслитель делает предположения, а потом избавляется от них. В конкретном случае Мыслитель начинает так: «Допустим, что он – Плут. Тогда...» Ну, вообще-то, это вопрос удобства, не необходимость. Я мог бы привести аргумент Мыслителя в следующей, более прямой форме: «Если он – Плут, тогда его утверждение ложно. Если его утверждение ложно, тогда я правильно верю, что он – Рыцарь. Если я правильно верю, что он – Рыцарь, тогда он – Рыцарь. Складывая все три эти факта вместе, получаем, что если он – Плут, тогда он – Рыцарь. Из этого последнего факта логически следует, что он – Рыцарь».

Общей практикой в логике является доказательство предложения формы $p \supset q$ (если p, тогда q) следующим образом: предполагается, что p истинно, и затем делается попытка вывести q. Если это может быть сделано, тогда предложение $p \supset q$ установлено. Другими словами, если предположение p в качестве посылки ведет к q как заключению, тогда $p \supset q$ доказано. (Эта техника есть часть того, что называется *натуральной дедукцией.*) Суть состоит в том, что все, что может быть доказано, используя эту технику, может быть также доказано без нее. (Есть хорошо известная теорема в логике на этот счет; она называется *теоремой дедукции*). И поэтому мы дозволим нашим «мыслителям» использовать натуральную дедукцию; Мыслитель типа 1, который делает это, не может доказать больше фактов, чем он может без использования натуральной дедукции, но доказательства с использованием натуральной дедукции значительно короче и легче для восприятия. Следовательно, мы будем позволять Мыслителю использовать натуральную дедукцию.

2. Дуальная проблема

Мы продолжаем предполагать, что Мыслитель принадлежит к типу 1, и что он верит в правила Острова, и что он слышит обращенные к нему утверждения.

Предположим, Обитатель, вместо того чтобы утверждать: «Ты никогда не поверишь правильно, что я – Рыцарь», утверждает: «Ты правильно поверишь, что я – Плут».

Получим ли мы при этом противоречие? (Читатель должен постараться решить эту проблему перед тем, как прочесть решение.)

Решение: Мыслитель рассуждает так: «Предположим, что он – Рыцарь. Тогда его утверждение истинно. Это означает, что я буду правильно верить, что он Плут, что в свою очередь влечет, что он – Плут. Отсюда, предположение, что он – Рыцарь, ведет к противоречию, следовательно, он должен быть Плутом».

На этом этапе Мыслитель верит, что Обитатель есть Плут, и он мыслит правильно, отсюда, Обитатель есть Плут. С другой стороны, так как Мыслитель верит правильно, что Обитатель есть Плут, утверждение Обитателя истинно, что делает его Рыцарем. Поэтому мы и в самом деле получаем противоречие.

Некоторые сходные проблемы

Давайте на некоторое время оставим Остров Рыцарей и Плутов и рассмотрим проблему, относящуюся к парадоксу в главе 3. Студент спрашивает профессора: «Существует ли на самом деле Бог?» И профессор дает следующий любопытный ответ: «Бог существует, если и только если ты неправильно веришь, что Он существует».

3.

Предположим, что студент есть Мыслитель типа 1, что утверждение профессора истинно и студент верит в него. Получаем ли мы при этом парадокс?

Решение: Да, мы получаем его! Начать с того, даже если забыть, что студент есть Мыслитель типа 1 и что он верит утверждению профессора, мы имеем следствие, что Бог должен существовать, потому что, если бы Бог не существовал, тогда студент *должен* был бы правильно верить, что Бог существует, но никто не может *правильно* верить в ложное утверждение. Следовательно, Бог должен реально существовать (в предположении, что утверждение профессора истинно).

Итак, студент есть Мыслитель типа 1 и поэтому знает пропозициональную логику, как вы и я, отсюда, он способен сделать заключение, что если утверждение профессора истинно, тогда Бог существует. Но он также верит утверждению профессора, и следовательно, он должен верить в существование Бога. И поскольку мы доказали, что Бог существует при трех предположениях, студент *правильно* верит, что Бог существует. Но Бог существует, если и только если студент *не* верит правильно в то, что Бог существует. Отсюда следует, что Бог не существует, если и только если студент правильно верит, что Бог существует. (Для любых предложений p и q предложение $p \equiv \neg q$ логически эквивалентно предложению $\neg p \equiv q$.) Поскольку Бог не существует, если и только если, студент правильно верит, что Бог существует, а студент правильно верит, что Бог существует, тогда следует заключение, что Бог не существует. Таким образом, три предположения, положенные в основу анализа проблемы, ведут к парадоксу, что Бог существует и Бог не существует.

Конечно, тот же самый парадокс возникнет, если профессор вместо сказанного до этого скажет: «Бог существует, если и только если ты правильно веришь, что Он не существует». Мы оставляем доказательство этого читателю в качестве упражнения.

Для нас сейчас существенно понять, что приведенный ваше парадокс является тем же самым, что и парадокс в Проблеме 1, относящийся к Острову Рыцарей и Плутов, хотя они могут показаться различными. Кажущиеся различия таковы: (1) в парадоксе выше профессор делает утверждение формы «если и только если», в то время как в Проблеме 1 Обитатель прямо говорит, что Мыслитель никогда не поверит правильно, что Обитатель есть Рыцарь. (2) В парадоксе студент верит профессору, в то время как в Проблеме 1 Мыслитель не имеет исходной веры, что Обитатель есть Рыцарь. Однако два этих различия в определенном смысле устраняют друг друга, как мы сейчас убедимся. Ключ к этому – устройство перевода из главы 7.

Практически во всех проблемах, которые будут обсуждаться позже, мы будем иметь дело только с двумя индивидами – Обитателем Острова, который делает утверждение, и Мыслителем, который слышит утверждение. Мы будем постоянно использовать букву k для предложения, что Обитатель есть Рыцарь. Тогда, как мы видели в главе 7, всякий раз, когда Обитатель утверждает предложение p, предложение k≡p будет истинным.

Далее, Мыслитель *верит* в правила Острова (он верит, что Рыцарь делает истинные утверждения, а Плут – ложные), и мы предполагаем, что он слышит утверждения, обращенные к нему. Следовательно, всякий раз, когда Обитатель утверждает предложение q, Мыслитель *верит* в предложение k≡q. В самом деле, начиная с этого момента, когда мы говорим, что соблюдаются правила Острова, мы имеем в виду не больше чем то, что для любого предложения q, если Обитатель утверждает q, тогда предложение k≡q истинно. И когда мы говорим, что Мыслитель верит в правила Острова (и слышит все обращенные к нему утверждения), при этом имеется в виду не больше чем то, что для любого предложения q, если Обитатель утверждает q, Мыслитель верит в предложение k≡q.

Для любого предложения p, Bp будет предложением, что Мыслитель верит (или будет верить) в p. И пусть C будет предложением p&Bp. Тогда Cp читается как «Мыслитель *правильно* верит в p».

В Проблеме 1 Обитатель утверждает предложение $\neg Ck$ («Ты никогда правильно не поверишь, что я Рыцарь». Так как правила Острова справедливы, предложение $k \equiv \neg Ck$ истинно. Поскольку Мыслитель верит в правила Острова, он верит в предложение $k \equiv \neg Ck$. Эти два факта логически несовместимы, и по этой причине возникает парадокс.

В Проблеме 3 пусть g будет предложением, что Бог существует. Тогда профессор говорит: $g \equiv \neg Cg$. При предположении, что профессор делает только истинные утверждения (Мыслитель), предложение $g \equiv \neg Cg$ должно быть истинным. Так как студент верит профессору, он верит в предложение $g \equiv \neg Cg$. Опять-таки истинность $g \equiv \neg Cg$ оказывается несовместимой с верой $g \equiv \neg Cg$ (так как студент есть Мыслитель типа 1).

Мы сейчас видим ясно, что есть общего в двух парадоксах: в обоих случаях мы имеем предложение p (k и g соответственно) такое, что $p \equiv \neg Cp$ одновременно истинно, и в него верит Мыслитель. Другими словами, в него правильно верит Мыслитель, а это логически невозможно, если Мыслитель типа 1. Таким образом, оба парадокса являются специальными случаями следующей теоремы:

Теорема А. Не существует предложения p такого, что Мыслитель типа 1 мог правильно верить в предложение $p \equiv \neg Cp$. Другими словами, не существует предложения p такого, что Мыслитель типа 1 мог правильно верить: «Предложение истинно, если и только если я не правильно верю, что оно истинно».

Доказательство Теоремы А представляет собой немного большее, чем просто повторение двух специальных случаев, упомянутых выше, но она может помочь рассмотреть это в более общей постановке и указать на некоторые интересные моменты.

Начать с того, что для любых предложений p и q предложение $(p\equiv\neg(p\&q))\supset p$ – тавтология (как это может проверить читатель). В частности, тавтологией будет $(p\equiv\neg(p\&Bp))\supset p$. Пусть $p\&Bp$ есть Cp, и поэтому $(p\equiv\neg Cp)\supset p$ есть тавтология. Предположим, что Мыслитель типа 1 правильно верит $p\equiv\neg Cp$. Тогда мы получим следующее противоречие: Так как Мыслитель *правильно* верит $p\equiv\neg Cp$, тогда $p\equiv\neg Cp$ должно быть истинным. Также $(p\equiv\neg Cp)\supset p$ истинно (это есть тавтология), поэтому p должно быть истинным. Так как Мыслитель типа 1, он *верит* в тавтологию $(p\equiv\neg Cp)\supset p$, и он также верит в $p\equiv\neg Cp$ (по предположению), и так как Мыслитель типа 1, он будет верить в p. И поэтому p истинно, и он верит p, поэтому он правильно верит в p. Таким образом, Cp истинно, отсюда, $\neg Cp$ ложно. Но так как p истинно, а $\neg Cp$ ложно, не может быть, чтобы $p\equiv\neg Cp$ (так как истинное предложение не может быть эквивалентно ложному предложению), и поэтому мы получаем противоречие, исходя из предположения, что Мыслитель типа 1 правильно верит в $p\equiv\neg Cp$.

Имеется следующая «дуальная» версия Теорема A, чье доказательство мы оставляем читателю.

Теорема A*. Не существует предложения p такого, что Мыслитель типа 1 может правильно верить в $p\equiv C(\neg p)$.

Упражнение 1. Доказать Теорему A*.

Упражнение 2. Предположим, что мы имеем полностью произвольную операцию B, которая приписывает каждому предложению p определенное предложение Bp. (Нет нужды в уточнении того, что представляет собой предложение Bp; в этой главе мы позволим Bp быть предложением таким, в которое *верит* Мыслитель; в последующей главе, в которой мы будем обсуждать не Мыслителей, а математические системы, Bp будет предложением, которое *доказуемо* в системе. Но пока Bp не уточняется.) Пусть Cp будет предложением (p&Bp).

(a) Показать, что можно вывести логическое противоречие из следующих предпосылок:

(i) Все предложения формы BX, где X – тавтология.

(ii) Все предложения формы $(BX\&B(X\supset Y))\supset BY$.

(iii) Некоторое предложение формы $C(p\equiv\neg Cp)$.

(b) Показать, что мы также получаем логическое противоречие, если мы заменим (iii) предложением «Некоторое предложение формы $C(p\equiv C\neg p)$».

(c) Почему Теорема A есть специальный случай (a) выше? Почему Теорема A* есть специальный случай (b)?

ГЛАВА 10. *Проблема усугубляется*

Самонадеянные Мыслители

Мы опять возвращаемся на Остров Рыцарей и Плутов. Предположим сейчас, что Обитатель вместо утверждения: «Ты никогда не поверишь правильно, что я есть Рыцарь» – делает следующее утверждение: *«Ты никогда не поверишь, что я есть Рыцарь»*.

Обитатель опустил слово «правильно», и в результате вещи становятся куда более интересными. Мы продолжаем предполагать, что Обитатель говорит Мыслителю типа 1, и что Мыслитель верит в правила Острова (а также то, что он слышит утверждения), и что правила соблюдаются. И сейчас мы сделаем дальнейшее предположение, что Мыслитель полностью точен в своих суждениях; он не верит в ложные предложения. Создает ли такая ситуация парадокс?

Предположим, что Обитатель – Плут. Тогда его утверждение ложно, что означает, что Мыслитель поверит, что Обитатель – Рыцарь. И поскольку Мыслитель точен в своих суждениях, Обитатель действительно Рыцарь. Таким образом, предположение о том, что Обитатель – Плут, ведет к заключению, что Обитатель – Рыцарь. Таким образом, предположение о том, что Обитатель – Плут, ведет к противоречию. Поэтому Обитатель должен быть Рыцарем.

Мыслитель принадлежит к типу 1 и знает столько из логики, как вы и я. Как предотвратить его от прохождения через тот же самый процесс размышления, через который мы прошли только что, и получения того же самого заключения, а именно, что Обитатель должен быть Рыцарем? Именно того, что Мыслитель будет верить, что Обитатель есть Рыцарь, и что делает утверждение Обитателя ложным, и отсюда, того, что Обитатель должен быть Плутом. Но мы уже доказали, что Обитатель есть Рыцарь. Парадокс!

1.

Изложенный выше аргумент ошибочен! Может ли читатель нащупать, в чем заключается ошибка? (Намек: Вопреки тому факту, что Мыслитель знает пропозициональную логику, так же как вы и я, имеется нечто, что знаем мы и не знает он. Что это?)

Решение: Я сказал вам, что Мыслитель всегда точен; но я никогда не говорил вам, что он *знает*, что он точен. Если бы он знал, что он точен (а этого он не мог знать), мы получили бы парадокс. Вы видите, что часть нашего доказательства, что Обитатель есть Рыцарь, использовала предположение, что Мыслитель всегда точен; если Мыслитель сделал это предположение, тогда он может подобным образом доказать, что Обитатель есть Рыцарь, таким образом сделав Обитателя Плутом.

Давайте отзовем предположение о том, что Мыслитель всегда точен в своих суждениях, но давайте предположим, что Мыслитель *верит* в то, что он всегда точен. Таким образом, для любого предложения p Мыслитель верит, что если бы он верил в p, тогда p должно быть истинным. Такой Мыслитель называется *самонадеянным* Мыслителем. Таким образом, самонадеянный Мыслитель это тот, который верит, что он неспособен верить ни в одно ложное предложение.

И поэтому мы отзываем посылку, что Мыслитель всегда точен, и заменяем посылкой, что Мыслитель верит в то, что всегда точен. Получаем ли мы тогда парадокс? Нет, вместо этого мы получаем следующий результат:

Теорема 1. Предположим, что Обитатель говорит Мыслителю типа 1: «Ты никогда не поверишь, что я – Рыцарь». Тогда, если Мыслитель верит, что всегда точен, он впадет в неточность, то есть рано или поздно он поверит во что-то ложное.

2.

Докажите Теорему 1.

Решение: Мыслитель рассуждает так: «Предположим, что Обитатель – Плут. Тогда его утверждение ложно, что означает, что я буду верить, что он – Рыцарь. Но если я когда-либо поверю, что он – Рыцарь, он должен быть таковым, потому что я не способен делать ошибок (заметьте!). Поэтому, если он – Плут, тогда он – Рыцарь, что невозможно. Следовательно, он Рыцарь».

На этом этапе Мыслитель верит в то, что Обитатель – Рыцарь. Так как Обитатель сказал, что Мыслитель никогда не поверит в это, тогда Обитатель на самом деле Плут. Поэтому Мыслитель сейчас имеет *ложную* веру в то, что Обитатель есть Рыцарь.

Интересно то, что если бы Мыслитель был более скромен и не предполагал собственной непогрешимости, он никогда бы не впал в неточность верования о том, что Обитатель есть Рыцарь. Мыслитель был просто наказан за свою самонадеянность!

Странные мыслители

Назовем Мыслителя *точным* в отношении данного предложения р, если вера Мыслителя в р влечет истинность р. То есть либо он не верит в р, либо он верит в р и р истинно. Мыслитель неточен, если он верит в р, и р ложно.

Важным фактом в отношении последней проблемы является то, что Мыслитель был неточен в отношении каждого предложения, в отношении которого он верил, что он точен, а именно в отношении предложения, что Обитатель – Рыцарь. Веря, что он точен в отношении этого предложения, он пришел к заключению, что Обитатель – Рыцарь, сделав, таким образом, обитателя Плутом. Конечно, наше доказательство, что Обитатель – Плут, основывается на предположении, что соблюдаются правила Острова. Предположим, что мы отзываем эту посылку, но продолжаем предполагать, что Мыслитель *верит,* что соблюдает правила Острова. Следует ли опять, что Мыслитель будет верить некоторым ложным предложениям? Конечно, Мыслитель все еще будет верить, что Обитатель – Рыцарь, хотя Обитатель сказал, что Мыслитель никогда не будет верить в это; но если правила острова не соблюдаются, тогда Обитатель не необходимо есть Плут. Однако, несмотря на то, что доказательство нашей Теоремы 1 рушится, если мы отзовем посылку о соблюдении правил Острова, мы имеем следующее озадачивающее утверждение:

Теорема 2. Предположим, что Обитатель говорит Мыслителю типа 1: «Ты никогда не будешь верить, что я – Рыцарь». Предположим, что Мыслитель *верит,* что соблюдаются правила Острова. Тогда независимо от того, соблюдаются эти правила или нет, если Мыслитель самонадеян, он придет к вере в некоторое ложное предложение.

3.

При посылках Теоремы 2 в какое ложное предложение будет верить Мыслитель?

Решение: Используя тот же самый метод доказательства, который использовался в Теореме 1, Мыслитель придет к вере, что Обитатель – Рыцарь. Эта вера не необходимо ложна (так как правила Острова не необходимо соблюдаются), но тогда Мыслитель продолжает: «Так как он – Рыцарь, и он сказал, что я никогда не поверю, что он – Рыцарь, тогда сказанное им должно быть истиной, то есть, я не верю (то есть не тот случай, что я верю), что он – Рыцарь. Поэтому я не верю, что он – Рыцарь».

На этом этапе Мыслитель верит, что Обитатель – Рыцарь, и также верит, что не верит, что обитатель – Рыцарь.

Поэтому он имеет ложную веру, что он не верит в то, что Обитатель – Рыцарь (эта вера ложна, потому что он *всетаки* верит, что обитатель – Рыцарь!

Заключение всего этого обсуждения весьма странно. Мыслитель верит, что Обитатель – Рыцарь, и также верит, что не верит, что Обитатель – Рыцарь. Но это не есть проявление противоречивости Мыслителя, хотя и есть в этом некоторая психологическая странность. Мы назовем Мыслителя *странным*, если имеется некоторое предложение p такое, что он верит в p и верит, что не верит в p. Это условие, конечно, влечет, что Мыслитель неточен (потому что он верит в ложное предложение, что он не верит в p). И по этой причине странный Мыслитель автоматически не точен, но не необходимо противоречив.

Мы будем говорить, что Мыслитель является *странным в отношении данного предложения p*, если он верит в p и верит, что не верит в p. Мыслитель тогда странен, если и только если имеется, по крайней мере, одно предложение p, в отношении которого он странен. Если Мыслитель странен в отношении p, тогда он не точен, но в отношении Bp, но не необходимо в отношении p.

Мы сейчас видим, что если мы отзовем посылку о соблюдении правил Острова, если Мыслитель верит, что они соблюдаются, и он Мыслитель типа 1, и Обитатель говорит ему, что Мыслитель никогда не поверит, что Обитатель – Рыцарь, тогда, если Мыслитель верит, что он точен в отношении предложения, что Обитатель – Рыцарь, его вера вынуждает его быть неточным в отношении предложения, что он *верит,* что Обитатель – Рыцарь. Если правила соблюдаются на самом деле, тогда Мыслитель будет также неточен в отношении предложения, что Обитатель – Рыцарь, то есть он будет верить, что обитатель – Рыцарь, в то время как на самом деле обитатель – Плут.

Конечно, та же самая проблема может быть сформулирована для задачи о существовании Бога. Предположим, что студент есть Мыслитель типа 1 и профессор говорит ему: «Бог существует, если и только если ты никогда не поверишь, что Он существует». Предположим также, что студент верит, что если он когда либо поверит в существование Бога, тогда Бог существует. Все это означает: (1) Если студент верит профессору, тогда он решает верить, что Бог существует, и также верить, что он не верит, что Бог существует. (2) Если утверждение профессора также истинно, тогда Бог не существует, но студент будет верить, что Бог все-таки существует.

Обе версии этой проблемы являются специальными случаями следующей теоремы:

Теорема А. Предположим, что Мыслитель принадлежит к типу 1, и что имеется предложение p такое, что он верит в предложение $p \equiv \neg Bp$, а также верит, что $Bp \supset p$. Тогда следует, что:

(а) Он будет верить в p, а также верить, что он не верит в p (он странен в отношении p).

(б) Если также $p \equiv \neg Bp$ истинно, тогда p ложно, но он будет верить в p.

Решение: (а) Он верит $p \equiv \neg Bp$, отсюда, он верит $Bp \supset \neg p$ (которое есть логическое следствие $p \equiv \neg Bp$). Он также верит в $Bp \supset p$ (по гипотезе), отсюда, он должен верить в $\neg Bp$ (которое есть логическое следствие $Bp \supset \neg p$ и $Bp \supset p$).

Но он также верит в $p \equiv \neg Bp$, отсюда, он должен верить в p (которое есть следствие $\neg Bp$ и $p \equiv \neg Bp$). И поэтому он также верит в p и также верит в $\neg Bp$ (он верит, что не верит в p!).

(б) Предположим, что $p \equiv \neg Bp$ истинно. Так как $\neg Bp$ ложно (он все-таки верит в p!), тогда p, будучи эквивалентным ложному предложению $\neg Bp$, также ложно. Следовательно, он верит в p, но p ложно.

Упражнение. Предположим, что Обитатель говорит самонадеянному Мыслителю типа 1: «Ты поверишь, что я – Плут». Доказать, что: (а) Если Мыслитель верит в правила Острова, тогда он верит, что Обитатель – Плут, и также не верит, что Обитатель – Плут. (б) Если при этом справедливы правила Острова, тогда Обитатель на самом деле – Рыцарь.

Мыслитель типа 1*

Под Мыслителем типа 1* мы имеем в виду Мыслителя типа 1 с дополнительным свойством, что для любых двух предложений p и q, если он вообще верит в предложение $p \supset q$, тогда он будет верить, что если он когда-либо поверит в p, тогда он также поверит в q. Символически, если Мыслитель когда-либо поверит в $p \supset q$, тогда он поверит в $Bp \supset Bq$.

Давайте заметим, что если Мыслитель типа 1 действительно верит в $p \supset q$, тогда истинно, что если он когда-либо поверит в p, тогда он поверит в q, то есть, $Bp \supset Bq$ есть истинное предложение (если он верит в $p \supset q$). От Мыслителя типа 1 этот Мыслитель типа 1* отличается тем, что если он верит в $p \supset q$, тогда не только $Bp \supset Bq$ истинно, но и то, что он правильно *верит* в $Bp \supset Bq$. Таким образом, Мыслитель типа 1* имеет большее «самосознание», чем Мыслитель типа 1.

Мы продолжаем предполагать, что Мыслитель, который принадлежит типу 1*, верит в правила Острова и слышит утверждения, ему адресованные. И всякий раз, когда Мыслитель утверждает p, Мыслитель верит в $k \equiv p$, где k – предложение, что Обитатель – Рыцарь.

Следующий факт будет чрезвычайно важен:

Лемма 1[1]**.** Предположим, что Обитатель делает утверждение Мыслителю типа 1*. Тогда Мыслитель будет верить, что если он когда-либо поверит, что Обитатель – Рыцарь, он также поверит тому, что Обитатель сказал.

Проблема 4. Как эта лемма доказывается?

Решение: (Доказательство чрезвычайно просто!) Предположим, что Обитатель говорит Мыслителю предложение p. Тогда Мыслитель верит предложению $k \equiv p$. Тогда он также верит $k \supset p$, поскольку это выражение есть логическое следствие $k \equiv p$. Тогда, поскольку он типа 1*, он верит $Bk \supset Bp$.

---

[1] Лемма есть предложение, доказываемое не столько ради него самого, сколько в качестве вспомогательного средства для доказательства последующих теорем. Вы можете сказать, что лемма есть предложение, которое недостаточно «благородно», чтобы называться теоремой.

Я назвал Мыслителя «самонадеянным», если он верит в свою собственную непогрешимость. Я не думаю, что вера человека в то, что он не странен, равносильна акту самонадеянности. В самом деле, предположение, что личность не является странной, является вполне разумным. Я даже не уверен, что психологически возможно для человека быть странным. Может ли человек на самом деле верить во что-то, а также верить, что он не верит в это? Я сомневаюсь в этом. И все же не является *логически* невозможным для человека быть странным.

В любом случае убежденность в собственной нестранности гораздо более разумна, чем убежденность в собственной полной точности. Так что следующая теорема отчасти печальна.

Теорема 3. Предположим, что Обитатель говорит Мыслителю типа 1*: «Ты никогда не поверишь, что я – Рыцарь». Тогда если Мыслитель верит, что он не является (и никогда не будет) странным, он будет странным!

Проблема 5. Доказать Теорему 3.

Решение: Предположим, что Обитатель делает это утверждение и что Мыслитель верит, что он неспособен быть странным. Так как Обитатель делает утверждение, тогда по Лемме 1, Мыслитель будет верить, что если он когда-либо поверит, что Обитатель – Рыцарь, он также поверит тому, что сказал Обитатель. И поэтому Мыслитель рассуждает так: «Предположим, если я поверю, что он – Рыцарь, тогда я поверю в то, что он сказал, то есть я поверю, что я не верю, что он – Рыцарь. И поэтому я поверю, что он – Рыцарь, и я также поверю, что я не верю, что он – Рыцарь. Это означает, что я буду странным. Следовательно, если я когда-либо поверю, что он – Рыцарь, я стану странным. Так как я никогда не буду странным (заметьте!), я никогда не буду верить, что он – Рыцарь. Так как он сказал, что я не буду верить, его утверждение истинно, и он – Рыцарь».

На этом этапе Мыслитель приходит к заключению, что Обитатель – Рыцарь, а ранее он пришел к заключению, что он не верит, что Обитатель – Рыцарь. Таким образом, он впал в странность.

Все это может быть установлено в более общей форме:

Теорема В. Для каждого Мыслителя типа 1*, если он верит в некоторое предложение формы $p \equiv \neg Bp$ («p истинно, если и только если, я никогда не поверю в p»), тогда он не может верить, что он не-странен, пока на самом деле не впадет в странность.

Мораль. Если вы – Мыслитель типа 1* и вы желаете верить, что вы не странны, вы можете избежать странности просто отказом верить в некоторое предложение формы «p, если и только если я никогда не поверю p».

В частности, если вы посещаете Остров и Обитатель говорит вам, что вы никогда не поверите, что он – Рыцарь, тогда самый мудрый курс – отказаться верить, что правила Острова соблюдаются.

Однако позднее в этой книге, когда мы подойдем к изучению математических систем и разговору о доказуемости в системе вместо вер Мыслителя, мы увидим, что аналог опции – отказ верить в $p \equiv \neg Bp$ – не будет доступен.

ЧАСТЬ V

ЗАТРУДНЕНИЯ С НЕПРОТИВОРЕЧИВОСТЬЮ

ГЛАВА 11. *Логики, которые мыслят о самих себе*

Мы все ближе подходим к геделевскому затруднению с непротиворечивостью. Но сначала нам требуется рассмотреть Мыслителей с более высокими степенями самосознания, чем просто тип 1.

Повышение стадий самосознания

Сейчас мы определим Мыслителей типов 2, 3, 4, которые представляют повышение степеней самосознания. Мыслитель типа 4 будет играть главную роль в развертывающейся драме.

Мыслитель типа 2. Предположим, что Мыслитель типа 1 верит в p и верит в $p \supset q$. Тогда он верит в q. Это означает, что предложение $(Bp \& B(p \supset q)) \supset Bq$ *истинно* для

Мыслителя типа 1. Однако Мыслитель не необходимо *знает,* что это предложение истинно. Мыслитель типа 2 есть мыслитель типа 1, который к тому же *верит* во все предложения формы $(Bp\&(p\supset q))\supset Bq$. (Он «знает», что его множество вер – прошлое, настоящее и будущее – замкнуто относительно *modus ponens*. Для любых предложений p и q он верит: «Если я верю в p и верю в $p\supset q$, тогда я верю в q»).

Стоит заметить суть отличия Мыслителей типов 1 и 2. Мыслитель типа 2 имеет определенное «самосознание», не необходимо присутствующее в Мыслителе типа 1. Мыслитель типа 1, который верит в p и верит в $p\supset q$, рано или поздно поверит в q. Мыслитель типа 2 также *знает*, что если он поверит в p и $p\supset q$, он также будет верить в q.

Мыслитель типа 3. Мы будем говорить, что Мыслитель *нормален*, если для любого предложения p, если он верит в p, тогда он верит, что верит в p (если он верит в p, тогда он верит в Bp). Под Мыслителем типа 3 мы имеем в виду нормального Мыслителя типа 2.

Мыслитель типа 3 имеет большую степень самосознания по сравнению с Мыслителем типа 2.

Мыслитель типа 4. Нормальный Мыслитель не необходимо знает, что он нормален. Если Мыслитель нормален, тогда для любого предложения p, предложение $Bp\supset BBp$ истинно, но Мыслитель не обязательно осознает истинность $Bp\supset BBp$. Мы говорим, что Мыслитель *верит*, что он нормален, если для каждого p, он верит в $Bp\supset BBp$. (Для каждого предложения p, Мыслитель верит: «Если я когда-либо поверю в p, тогда я поверю, что я верю в p.)

Мыслитель типа 3 на самом деле нормален. Под Мыслителем типа 4 мы имеем в виду Мыслителя типа 3, который к тому же *знает,* что он нормален. Таким образом, для любого предложения p Мыслитель типа 4 верит в $Bp \supset BBp$.

Как мы уже замечали, Мыслитель типа 4 играет главную роль в этой книге. Давайте рассмотрим условия, определяющие Мыслителя типа 4:

(1а) Он верит во все тавтологии;
(1б) Если он верит в p и $p \supset q$, тогда он верит в q;
(2) Он верит в $(Bp \& B(p \supset q)) \supset Bq$;
(3) Если он верит в p, тогда он верит в Bp;
(4) Он верит в $Bp \supset BBp$.

Некоторые основные свойства самосознающих Мыслителей

Мы сейчас установим несколько простых свойств Мыслителей типов 2, 3 и 4, которые будут использованы в дальнейших главах.

Во-первых, это простое наблюдение о Мыслителе типа 2: Для любых p и q предложение $(Bp \& B(p \supset q)) \supset Bq$ логически эквивалентно предложению $B(p \supset q) \supset (Bp \supset Bq)$, потому что для любых X, Y, Z предложение $(X \& Y) \supset Z$ эквивалентно $Y \supset (X \supset Z)$, что легко проверить, – и поэтому любой Мыслитель типа 2 верит во все предложения $B(p \supset q) \supset (Bp \supset Bq)$. Обратно, любой Мыслитель типа 1, который верит во все предложения формы $B(p\supset q)\supset(Bp\supset Bq)$, должен быть Мыслителем типа 2. Давайте назовем это Фактом 1.

Факт 1. Мыслитель типа 1 есть Мыслитель типа 2, если и только если он верит во все предложения $B(p\supset q)\supset(Bp\supset Bq)$.

Предположим далее, что Мыслитель типа 2 верит в $B(p\supset q)$. Он также согласно, Факту 1, верит в $B(p\supset q)\supset(Bp\supset Bq)$, и будучи Мыслителем типа 1 (так как он типа 2), он также верит в $Bp\supset Bq$ (как логическое следствие двух предыдущих). И в качестве явного следствия Факта 1 мы имеем следующее утверждение: Если Мыслитель типа 2 верит в $B(p\supset q)$, тогда он верит в $Bp\supset Bq$.

Теперь мы подходим к менее очевидным фактам относительно Мыслителя типа 2.

1.

Покажите, что для любого Мыслителя типа 2 и любых трех предложений p, q и r:

(а) Он будет верить в $B(p\supset(q\supset r))\supset(Bp\supset(Bq\supset Br))$.

(б) Если он когда-либо будет верить в $B(p\supset(q\supset r))$, тогда он будет верить в $(Bp\supset(Bq\supset Br))$.

Решение: Предположим, что Мыслитель принадлежит к типу 2. Возьмем любые предложения p, q, r. Согласно Факту 1 он верит в:

(1) $B(p\supset(q\supset r))\supset(Bp\supset(Bq\supset Br))$.

Причина, по которой он верит в это, состоит в том, для любых двух предложений X и Y, он верит в $B(X\supset Y)\supset\supset(BX\supset BY)$, так что берем p для X и $(p\supset q)$ для Y.

Вновь, по Факту 1, он верит в

(2) $B(q\supset r)\supset(Bq\supset Br)$.

Следующее предложение есть логическое следствие (2)

(3) $B(p\supset(q\supset r))\supset(Bp\supset(Bq\supset Br))$.

Причина этого состоит в том, что для любых предложений X,Y и Z, предложение $(X\supset Y)\supset(X\supset Z)$ есть логическое следствие предложения $Y\supset Z$, что легко проверить. Мы берем Bp для X, $B(q\supset r)$ для Y и $Bq\supset Br$ для Z. Мы видит при этом, что (3) есть логическое следствие (2).

Мы теперь знаем, что Мыслитель верит в (1) и (2) и $B(p\supset(q\supset r))\supset(Bp\supset(Bq\supset Br))$ есть логическое следствие (1) и (2) и поэтому Мыслитель верит в него. Это доказывает (a).

(b) Предположим, что мыслитель верит в $B(p\supset(q\supset r))$. Согласно (a) он также верит в $B(p\supset(q\supset r))\supset(Bp\supset(Bq\supset Br))$; отсюда, он верит в $(Bp\supset(Bq\supset Br))$, так как это есть логическое следствие двух предложений выше.

Замечание: Я не буду далее давать такие детальные объяснения. Сейчас читатель должен иметь достаточный опыт в следовании аргументам и может восстановить опущенные шаги.

2.

Покажите, что если Мыслитель типа 3 верит в $p\supset(q\supset r)$, тогда он будет верить в $(Bp\supset(Bq\supset Br))$. Этот факт будет иметь многие применения.

Доказательство: Рассмотрим сейчас Мыслителя типа 3, который верит в $p\supset(q\supset r)$. Так как он нормален, он верит в $B(p\supset(q\supset r))$. Тогда по (b) выше он поверит в $Bp\supset(Bq\supset Br)$.

3. Регулярность

В последней главе мы определили Мыслителя типа 1* как Мыслителя типа 1, для которого есть условие, а именно: если для любых предложений p и q, если он верит в $p\supset q$, он также будет верить в $Bp\supset Bq$. Этому второму условию будет дано имя – мы назовем Мыслителя *регулярным*, если его вера в $p\supset q$ влечет его веру в $Bp\supset Bq$.

Доказать, что каждый Мыслитель типа 3 является регулярным и, таким образом, типа 1*.

Решение. Предположим, Мыслитель типа 3 верит в $p\supset q$. Так как он нормален, он будет верить в $B(p\supset q)$. Тогда, вследствие Факта 1, он будет верить в $Bp\supset Bq$.

4.

Доказать, что если регулярный Мыслитель типа 1 верит в $p\equiv q$, тогда он будет также верить в $Bp\equiv Bq$.

Решение: Предположим, что регулярный Мыслитель типа 1 верит в $p \equiv q$. Тогда он будет верить в $p \supset q$ и $q \supset p$ (поскольку они являются следствиями $p \equiv q$). Будучи регулярным, он тогда будет верить в $Bp \supset Bq$ и $Bq \supset Bp$. Тогда, будучи Мыслителем типа 1, он также поверит в $Bp \equiv Bq$ (которое является следствием двух утверждений выше).

5.

Есть интересная связь между регулярностью и нормальностью. Для регулярного Мыслителя типа 1, если имеется такое предложение, как q, такое, что Мыслитель верит в Bq, тогда он должен быть нормальным. Почему?

Решение: Рассмотрим регулярного Мыслителя типа 1. Предположим, что q есть некоторое предложение такое, что Мыслитель верит в Bq. Пусть p будет некоторым предложением, в которое Мыслитель верит. Нам нужно тогда показать, что он также верит в Bp.

Предложение $p \supset (q \supset p)$ является тавтологией (читатель может проверить это), отсюда Мыслитель верит в него. Он также верит в p (по предположению), отсюда он верит в $q \supset p$. Так как он регулярен, он поверит в $Bq \supset Bp$. Тогда, так как он верит в Bq, он поверит в Bp. Это доказывает, что он нормален.

6.

Любой регулярный Мыслитель типа 1, который верит в p и верит в q, будет верить в p&q (которое есть логическое следствие двух предложений – p и q). Таким образом, предложение $(Bp\&Bq)\supset B(p\&q)$ *истинно* для любого Мыслителя типа 1, и отсюда, истинно для типа 3.

Доказать, что любой Мыслитель типа 3 *верит* в $(Bp\&Bq)\supset B(p\&q)$. (Он знает, что если он поверит в p и поверит в q, он будет верить в p&q.)

Решение: Мы рассматриваем Мыслителя типа 3. Предложение $p\supset(q\supset(p\&q))$ явно тавтология, и поэтому Мыслитель поверит в него. Тогда, согласно (b) из Проблемы 1, он поверит в $Bp\supset(Bq\supset B(p\&q))$, отсюда, он поверит логически эквивалентному предложению $(Bp\&Bq)\supset B(p\&q)$. (Для любых предложений X, Y, Z предложение $X\supset(Y\supset Z)$ логически эквивалентно предложению $(X\&Y)\supset Z$.)

Непротиворечивость

Мы будем говорить, что Мыслитель *непротиворечив*, если множество предложений, в которое он верит, является непротиворечивым множеством, и мы будем говорить, что он *противоречив*, если противоречиво множество его вер. Для Мыслителя типа 1 множество его вер логически замкнуто; отсюда из Принципа C (глава 8) следует, что три условия ниже эквивалентны:

(1) Он противоречив (он верит в $\perp$).
(2) Он верит в p и $\neg p$.
(3) Он верит во все предложения.

Мы будем говорить, что Мыслитель *верит,* что он непротиворечив, если он верит в $\neg B\bot$ (он верит, что не верит в $\bot$). Мы будем говорить, что он верит, что он противоречив, если он верит в $B\bot$ (он верит, что верит в $\bot$).

Мыслитель даже типа 1, который противоречив, будет также верить, что он противоречив (поскольку он верит во все), хотя Мыслитель, который верит, что он противоречив, не необходимо противоречив (хотя может быть показано, что он должен иметь по крайней мере одну ложную веру).

Любой Мыслитель типа 1, который верит в p и его отрицание, будет противоречив, и поэтому предложение $(Bp\&B\neg p)\supset B\bot$ *истинно* для Мыслителя типа 1.

7.

Доказать, что любой Мыслитель типа 3 верит в $(Bp\&$ $\&B\neg p)\supset B\bot$.

Решение: Предложение $(p\&\neg p)$ есть тавтология, отсюда, любой Мыслитель типа 3 (или даже типа 1) поверит в него. Так как Мыслитель типа 3 регулярен (по Проблеме 3), он будет также верить в $B(p\&\neg p)\supset B\bot$. Он также верит в $(Bp\&B\neg p)\supset B(p\&\neg p)$ (по Проблеме 6). Поверив в последние два предложения, он поверит в $((Bp\&B\neg p)\supset B\bot$, которое является их логическим следствием.

Замечание к решению. Для *любого* предложения q предложение $p\supset(\neg p\supset q)$ есть тавтология. Отсюда, согласно изложенному выше аргументу применительно к q, взамен $\perp$ Мыслитель типа 3 будет верить в $(Bp\&B\neg p)\supset Bq$.

Замечание. Эта последняя проблема является решающей для следующей главы. Это означает, что для любого p, Мыслитель типа 3 *знает*, что если он поверит в p, а также в $\neg p$, тогда он будет противоречив. (Конечно, это также применимо и к Мыслителю типа 4, поскольку каждый тип 4 также принадлежит к типу 3).

Противоречивость и странность

Мы говорили ранее, что Мыслитель называется странным, если он верит в некоторое предложение p и также верит, что он не верит в предложение p. Мы отметили, что Странный Мыслитель не является необходимо противоречивым. Однако некоторый странный Мыслитель типа 3 *является* противоречивым, как это видно из следующей проблемы.

8

Доказать, что некоторый странный нормальный Мыслитель типа 1 должен быть противоречивым (и отсюда должен быть противоречивым некоторый странный Мыслитель типа 3).

Решение: Это довольно просто. Если нормальный Мыслитель верит в p, тогда он верит в Bp. Если он также верит в $\neg Bp$ и принадлежит к типу 1, он будет также противоречивым. Таким образом, если нормальный Мыслитель типа 1 верит в p и верит в B¬p, он будет противоречив.

Упражнение 1. Согласно проблеме выше, для любого p, предложение (Bp&B¬Bp)⊃B⊥ *истинно* для Мыслителя типа 3, и отсюда, также истинно для типа 4. Доказать, что некоторый Мыслитель типа 4 правильно *верит* в предложение (Bp&B¬Bp)⊃B⊥ (он *знает*, что если он когда-либо будет странным, он будет противоречив).

Решение упражнения 1: Мыслитель типа 4 верит в Bp⊃BBp. Он также верит в его логическое следствие (Bp&B¬Bp)⊃(BBp&B¬Bp). Он также верит в (BBp&B¬¬Bp)⊃B⊥ (согласно Проблеме 7, так как ¬Bp есть отрицание Bp). Предложение (Bp&B¬p)⊃B⊥ есть логическое следствие двух последних предложений, и поэтому Мыслитель верит в него.

Другими словами, Мыслитель типа 4 может размышлять таким образом: «Предположим, что я поверю в p, а также поверю в B¬p. Так как я верю в p, я также верю в Bp, отсюда я верю в оба предложения Bp и B¬p, и тогда я буду противоречив. И поэтому, если я поверю в p и в ¬Bp, я буду противоречив».

9. Маленькая загадка

Предположим, что Мыслитель типа 4 верит в p≡Bq. Будет ли он необходимо верить в p⊃Bp?

Решение: Предположим, что Мыслитель типа 4 верит в $p \equiv Bq$. Он является регулярным (согласно Проблеме 3, так как он Мыслитель типа 3), и поэтому, согласно проблеме 3, он будет верить в $Bp \equiv BBq$. Отсюда, он будет верить в $BBq \supset Bp$. Он также верит в $Bq \supset BBq$ (так как он Мыслитель типа 4), и отсюда, по пропозициональной логике, он будет верить в $Bq \supset Bp$. Из $p \equiv Bp$ и $Bq \supset Bp$ он сделает вывод $p \supset Bp$. И тогда мы имеем утвердительный ответ на загадку.

Осознание самосознания

Мыслитель типа 4 имеет одно удивительное свойство, которое не разделяется его низшими аналогами, а именно, что *знает,* что он типа 4, знает в том смысле, который мы сейчас определим точно. Например, Мыслитель может быть типа 3, не зная этого, но Мыслитель не может быть типа 4, не зная об этом.

Давайте скажем, что Мыслитель *верит,* что он типа 1, если он верит во все предложения формы BX, где X – тавтология, и верит во все предложения формы $Bp \& B(p \supset q)) \supset \supset Bq$. Если он также *верит* во все предложения формы $B(Bp \& B(p \supset q)) \supset Bq)$, тогда мы будем говорить, что он верит, что он типа 2. Если он также верит во все предложения формы $Bp \supset BBp$, тогда мы будем говорить, что он *верит,* что он Мыслитель типа 3. Если он также верит во все предложения формы $B(Bp \supset BBp)$, тогда мы будем говорить, что он *верит,* что он – Мыслитель типа 4. Для каждого из этих типов мы будем говорить, что Мыслитель *знает*, какого он типа, если он верит, что он этого типа, и на самом деле этого типа.

Нетрудно видеть, что Мыслитель, который знает, что он типа 1, принадлежит к типу 2. Мыслитель, который знает, что он типа 2, принадлежит типу 3 (хотя не необходимо знает, что он типа 3). Мыслитель принадлежит типу 4, если и только если он знает, что он типа 3. Читатель должен попытаться доказать эти факты в качестве упражнений.

Следующая проблема более интересна.

10.

Доказать, что Мыслитель типа 4 знает, что он типа 4.

Эта проблема интересна по нескольким причинам. Во-первых, она показывает, что Мыслитель типа 4 является естественным конечным пунктом нашей иерархии Мыслителей. (Было бы бесполезно, например, определять Мыслителя типа 5, который знает, что он типа 4, так как тип 4 уже знает это о себе, и мы не получили бы ничего нового.)

Во-вторых, все, что вы можете доказать о Мыслителях типа 4, используя просто пропозициональную логику, любой Мыслитель типа 4 может доказать сам о себе, так как он знает пропозициональную логику и знает, что он – Мыслитель типа 4.

Решение: Предположим, что Мыслитель типа 4. Тогда он удовлетворяет всем условиям, определяющим Мыслителя типа 4. Мы должны показать, что он *верит* во все эти условия.

(1a) Возьмем любую тавтологию X. Будучи Мыслителем типа 4 (и отсюда, типа 1), он верит в X. Тогда, поскольку он нормальный Мыслитель, он верит в BX. Таким образом, для любой тавтологии X он верит в BX.

(1б) Из факта, что он Мыслитель типа 2, следует, что он верит во все предложения формы $(Bp\&B(p\supset q))\supset Bq$.

На этом этапе мы понимаем, что он знает, что он Мыслитель типа 1. (На самом деле, согласно изложенному выше аргументу, любой Нормальный Мыслитель типа 2 например любой Мыслитель типа 3, знает, что он – Мыслитель типа 1).

(2) Поскольку он верит во все предложения формы $(Bp\&B(p\supset q))\supset Bq$ и является Нормальным Мыслителем, он верит в $B(Bp\&B(p\supset q))\supset Bq)$.

На этом этапе мы видим, что он знает, что он – Мыслитель типа 2. (На самом деле, любой Мыслитель типа 3 знает, что он – Мыслитель типа 2).

(3). Так как он – Мыслитель типа 4, тогда непосредственно следует, что он знает все предложения формы $Bp\supset BBp$, то есть он знает, что он Нормален.

На этом этапе мы видим, что Мыслитель типа 4 знает, что он – Мыслитель типа 3. (Но Мыслитель типа 3 не необходимо знает, что он типа 3, потому что он может не знать, что он Нормален.)

(4). Так как Мыслитель типа 4 верит в $Bp \supset BBp$ и он Нормален, он верит в $B(Bp\supset BBp)$.

На этом этапе Мыслитель типа 4 знает, что он – Мыслитель типа 4 (то есть он знает все предложения, характеризующие Мыслителя типа 4).

Упражнение 2. Утверждать, что Мыслитель Регулярен, значит утверждать, что для любых предложений p и q утверждение $B(p\supset q)\supset B(Bp\supset Bq)$ *истинно* для Мыслителя. (Предложение $B(p\supset q)\supset B(Bp\supset Bq)$ есть предложение, что если Мыслитель верит в $p\supset q$, тогда он верит в $Bp\supset Bq$.) Давайте говорить, что Мыслитель *верит*, что он Регулярен, если он *верит* во все предложения формы $B(p\supset q)\supset$ $\supset B(Bp\supset Bq)$.

Доказать, что каждый Мыслитель типа 4 знает, что он Регулярен (то есть что он Регулярен и верит, что он Регулярен).

Решение к Упражнению 2: Рассмотрим Мыслителя типа 4. Как Мыслитель типа 1 он верит в $B(p\supset q)\supset(Bp\supset Bq)$. Так как он Регулярен, он верит в $BB(p\supset q)\supset B(Bp\supset Bq)$. Он также верит в $B(p\supset q)\supset BB(p\supset q)$, поскольку он нормален. Из этого последнего факта следует, что он должен верить в $B(p\supset q)\supset B(Bp\supset Bq)$.

Упражнение 3. Докажите, что если Мыслитель типа 4 верит в $p\supset(Bp\supset q)$, тогда он будет верить в $Bp\supset Bq$. (Решение этого упражнения будет дано в главе 15.)

ГЛАВА 12. *Затруднения с непротиворечивостью*

Теперь сцена готова и начинается настоящее шоу!

Мыслитель типа 4 посещает Остров Рыцарей и Плутов и верит в правила Острова. И поэтому всякий раз, когда Обитатель делает утверждение, Мыслитель будет верить, что если Обитатель есть Рыцарь, это утверждение истинно, и обратно. Таким образом, если Обитатель делает утверждение p, тогда Мыслитель верит в $k \supset p$ (где k есть утверждение Обитателя), и он также верит в $p \supset k$. Больше того, поскольку Мыслитель типа 4, он регулярен, и поэтому верит не только в $k \supset p$, но и в $Bk \supset Bp$, то есть он будет верить в утверждение: «Если я когда-либо поверю, что он – Рыцарь, тогда я поверю в сказанное им).

Мы вспомним из последней главы (Проблема 7), что любой Мыслитель типа 4 или даже типа 3 знает, что если он когда-либо поверит в p и поверит в $\neg p$, он будет противоречив (p может быть любым предложением).

Давайте рассмотрим все эти факты.

Факт 1. Пусть Обитатель делает утверждение Мыслителю типа 4. Тогда: (а) Мыслитель поверит, что если Обитатель есть Рыцарь, утверждение должно быть истинным, и обратно. (б) Мыслитель также поверит, что если он поверит, что Обитатель – Рыцарь, тогда он поверит в то, что говорит Обитатель.

Факт 2. Для любого предложения p, Мыслитель типа 4 знает, что если он поверит p и также поверит в $\neg p$, тогда он противоречив.

Имея в виду эти факты, мы готовы предпринять наше путешествие. Первым большим результатом, к которому мы сейчас обратимся, является следующая теорема.

Теорема 1. (По мотивам теоремы Геделя о непротиворечивости.) Предположим, что Обитатель говорит Мыслителю типа 4: «Ты никогда не поверишь, что я – Рыцарь». Тогда, если Мыслитель непротиворечив, он никогда не узнает, что он непротиворечив, или же, другими словами, если Мыслитель когда-либо поверит, что он не может быть противоречивым, он будет противоречивым!

1.

Доказать Теорему 1.

Решение: Предположим, что Мыслитель все же верит, что он непротиворечив (и будет оставаться таковым). Мы покажем, что он станет противоречивым.

Мыслитель рассуждает так: «Предположим, что я когда-либо поверю, что Обитатель есть Рыцарь. Тогда я буду верить в то, что он сказал, – я буду верить, что не верю, что он – Рыцарь. Кроме того, если я верю, что он – Рыцарь, тогда я буду верить, что я все же верю, что он – Рыцарь (так как я нормален). Следовательно, если я когда-либо поверю, что он – Рыцарь, тогда я буду верить одновременно, что я верю, что он Рыцарь, и что я не верю, что он – Рыцарь, что означает мою противоречивость. Но я никогда не буду противоречивым (заметьте!), отсюда, я никогда не буду верить, что он – Рыцарь. Он сказал, что я никогда не поверю, что он – Рыцарь, и то, что он сказал, было истинно, отсюда, он – Рыцарь».

На этом этапе Мыслитель верит, что Обитатель есть Рыцарь, и, поскольку Мыслитель нормален, он будет знать, что он верит в это. Поэтому Мыслитель будет продолжать рассуждать: «Сейчас я верю, что он – Рыцарь. Он сказал, что я никогда не поверю в это, поэтому он сделал ложное утверждение и поэтому он не Рыцарь».

На этом этапе Мыслитель верит, что Обитатель есть Рыцарь, и также верит, что Обитатель не есть Рыцарь, и поэтому Мыслитель противоречив.

Обсуждение. Важное математическое содержание этой теоремы может быть представлено без Мыслителя и Обитателя. Их функция состоит в том, чтобы обеспечить простой метод получения некоторого предложения k (в данном случае Обитатель есть Рыцарь) такого, что Мыслитель будет верить в «k истинно, если и только если, я никогда не поверю в k». Любой другой метод получения такого предложения k будет служить также хорошо. Таким образом, Теорема 1 является лишь специальным случаем следующей теоремы:

Теорема G. Если непротиворечивый Мыслитель типа 4 верит в некоторое предложение формы $p \equiv \neg Bp$, тогда Мыслитель никогда не может знать, что он непротиворечив.

Другими словами, если Мыслитель типа 4 верит в $p \equiv \neg Bp$ и верит в то, что он непротиворечив, тогда он становится противоречивым.

Мы докажем Теорему G в более точной форме:

Теорема G#. Предположим, что нормальный Мыслитель типа 1 верит в предложение формы $p \equiv \neg Bp$. Тогда:

(а) Если он когда-либо поверит в p, он станет противоречивым.
(б) Если он типа 4, тогда он знает, что если он когда-либо поверит в p, он станет противоречивым, то есть он будет верить, например, в предложение $Bp \supset B\bot$.
(в) Если он типа 4 и верит в то, что он не может быть противоречивым, тогда он станет противоречивым.

Доказательство: (а) Предположим, что он верит в p. Будучи нормальным, он тогда верит в Bp. Также, поскольку он верит в p и верит в $p \equiv \neg Bp$, он должен верить в $\neg Bp$ (так как он типа 1). И он будет тогда верить как в Bp, так и в $\neg Bp$, и отсюда он будет противоречивым.

(б) Предположим, что он типа 4. Так как он типа 1 и верит в $p \equiv \neg Bp$, он должен также верить в $p \supset \neg Bp$. Далее, поскольку он регулярен, он будет тогда верить в $Bp \supset (B\neg Bp)$. Кроме того, он также верит в $Bp \supset BBp$ (так как он знает, что нормален). Отсюда он будет верить в $Bp \supset (BBp \& B\neg Bp)$, которое является логическим следствием двух последних предложений. Он также верит в $(BBp \& B\neg Bp) \supset B\bot$ (по Факту 2, так как для любого предложения X он верит в

$(BX\&B\neg X)\supset B\bot$, и поэтому он верит в это в специальном случае, где X есть Bp). Раз он верит в оба предложения $Bp\supset(BBp\&B\neg Bp)$ и $(BBp\&B\neg Bp)\supset B\bot$, он будет также верить в $Bp\supset B\bot$ (так как он типа 1).

(в) Так как он верит в $Bp\supset B\bot$ (что мы только что доказали), он также верит в $\neg B\bot\supset\neg Bp$. Теперь предположим, что он верит в $\neg B\bot$ (он верит, что не может быть противоречивым). Так как он также верит в $\neg B\bot\supset\neg Bp$ (как мы только сейчас это видели), он поверит и в $\neg Bp$. Поскольку он также верит в $p\equiv\neg Bp$, он поверит в p и, отсюда, будет противоречив, согласно (а).

Студент и профессор теологии. Давайте снова вернемся к студенту и его профессору теологии, который говорит студенту: «Бог существует, если и только если, ты никогда не поверишь, что Он существует». Если студент верит профессору, тогда он верит в предложение $g\equiv\neg Bg$, где g – предложение «Бог существует». Тогда, по Теореме G, студент не может верить в свою собственную непротиворечивость, не становясь при этом противоречивым.

Мы намекнули на это в главе 2, но тогда мы не были способны установить, что составляет «разумное» множество предположений о мыслительных способностях студента. Сейчас мы можем сделать это. Предположение состоит просто в том, что студент является Мыслителем типа 4.

Конечно, студент имеет возможность верить в свою собственную непротиворечивость, не впадая при этом в противоречивость; он может просто не верить профессору!

Упражнение 1. Предположим, что в Теореме 1 нам дана дополнительная информация, что правила Острова на самом деле соблюдаются. Является при этом Обитатель Рыцарем или Плутом?

Ответ на Упражнение 1. По Теореме 1, Мыслитель будет противоречив, и поэтому он будет верить во что угодно. В частности, он будет верить в то, что Обитатель есть Рыцарь. Так как Обитатель говорит, что он не мог бы верить в это, утверждение Обитателя ложно. Следовательно, если правила Острова на самом деле соблюдаются, Обитатель должен быть Плутом.

Упражнение 2. В примере со студентом и профессором теологии предположим, что студент верит профессору, а также верит в свою собственную непротиворечивость. Если Бог реально существует, тогда утверждение профессора было истинным или ложным? Если Бог не существует, тогда утверждение профессора было истинным или ложным?

Ответ на Упражнение 2. Вновь студент будет противоречивым и верить во все, что угодно. В частности, он верит, что Бог существует (он будет также верить, что Бог не существует, но это в данном контексте несущественно). Пусть g будет предложение о том, что Бог существует, тогда Bg будет истинным, отсюда, $\neg Bg$ будет ложным. Если Бог существует, тогда g истинно, отсюда $g \equiv \neg Bg$ будет ложно, и профессор был неправ. Если Бог не существует, тогда g ложно, $g \equiv \neg Bg$ истинно, и профессор был прав.

Упражнение 3. Мы доказали в предшествующих главах следующие три факта:

(1) Если Обитатель говорит Мыслителю типа 1*: «Ты никогда не поверишь, что я есть Рыцарь», – и Мыслитель верит, что никогда не будет странным, тогда он становится странным (Теорема 3, глава 10).
(2) Странный Мыслитель типа 3 противоречив (Проблема 8, глава 11).
(3) Мыслитель типа 4 верит, что если он когда-либо станет странным, тогда он станет противоречивым (Упражнение 1, глава 11).

Используя эти три факта, можно дать гораздо более быстрое доказательство Теоремы 1 данной главы, чем приведенное выше. Как это сделать?

Ответ на Упражнение 3. Предположим, что Обитатель делает утверждение Мыслителю типа 4: «Ты никогда не поверишь, что я есть Рыцарь», – и Мыслитель при этом верит, что никогда не будет противоречивым. Тогда Мыслитель будет верить, что он не может быть странным (потому что он знает, что если он странен, он будет противоречивым). Тогда, согласно Факту 1, он становится странным. И, согласно Факту 2, он будет противоречив.

Дуальная к Теореме G

Упражнение 4. Предположим, что Обитатель, вместо утверждения «Ты никогда не поверишь, что я есть Рыцарь», говорит: «Ты *поверишь*, что я – Плут». Предположим, что Мыслитель принадлежит к типу 4 и верит в свою собственную непротиворечивость. Следует ли отсюда, что он при этом становится противоречивым?

Решение: Ответ утвердительный, и он может быть установлен как следствие Теоремы G, но сперва я хотел бы набросать прямое доказательство. Мыслитель рассуждает так: «Предположим, что он – Рыцарь. Тогда сказанное им истинно, что означает, что я поверю в то, что он есть Плут. Раз я поверил, что он есть Плут, я поверю в *противоположное* тому, что он сказал – я поверю, что *не* верю, что он есть Плут. Но если я верю, что он есть Плут, я также верю, что я *все-таки* верю, что он есть Плут (потому что я нормален), и отсюда, я становлюсь противоречивым. Так как я никогда не буду противоречивым, он не может быть Рыцарем, так что он должен быть Плутом. Итак, я верю, что он – Плут. Он сказал, что я буду верить в это, и поэтому он есть Рыцарь.

На этом этапе Мыслитель противоречив.

То, что мы сейчас доказали, есть специальный случай следующей дуальной к Теоремы G.

Теорема G°. Если непротиворечивый Мыслитель типа 4 верит в некоторое предложение формы $p \equiv B\neg p$, тогда он никогда не узнает, что он непротиворечив.

2.

Теорема G° может быть доказана дуальным к Теореме G аргументом, но она может быть доказана более простым путем как следствие Теоремы G. Как?

Решение: Предположим, что Мыслитель верит в $p \equiv B \neg p$. Тогда он будет верить также в $\neg p \equiv \neg B \neg p$. Пусть q будет предложением $\neg p$. Тогда Мыслитель верит в $q \equiv \neg Bq$, и поэтому Мыслитель верит также в предложение формы $p \equiv \neg Bp$ (а именно: $q \equiv \neg Bq$), и поэтому результат следует из Теоремы G.

Упражнение 5. Предположим, что в Упражнении 4 мы добавляем предположение, что правила Острова на самом деле выполняются, и что Мыслитель *все-таки* верит в свою собственную непротиворечивость. Кем тогда является Обитатель – Рыцарем или Плутом?

Упражнение 6. Предположим, что профессор теологии говорит студенту типа 4: «Бог существует, если и только если ты поверишь, что Бог не существует». Предположим, что студент верит профессору, а также верит в свою собственную непротиворечивость. Доказать, что если утверждение профессора истинно, тогда Бог должен существовать. Доказать, что если утверждение профессора ложно, тогда Бог не существует.

Упражнение 7. Предположим, что Обитатель говорит Мыслителю типа 4: «Ты никогда не поверишь, что я есть Рыцарь» (или альтернативно: «Ты поверишь, что я есть Плут»). Доказать, что Мыслитель знает, что если он когда-либо поверит в свою собственную непротиворечивость, он станет противоречивым. (Решение следует легко из результатов, которые будут доказаны в дальнейших главах).

ГЛАВА 13. *Геделевы системы*

Все доказанные до сих пор результаты о Мыслителях и их верах являются аналогами метаматематических результатов о математических системах и предложениях, доказуемых в них. Перед тем как обратиться именно к ним, давайте подведем итоги в отношении наиболее важных фактов, которые мы доказали в последних нескольких главах.

Краткие итоги I

Предположим, что Мыслитель верит в законы Острова, и Обитатель говорит ему: «Ты никогда не поверишь, что я – Рыцарь». Или, более обще, предположим, что Мыслитель верит в некоторое предложение формы $p \equiv \neg Bp$ (p истинно, если и только если я никогда не поверю в p). Тогда:

> (1) Мыслитель типа 1, если он верит, что всегда точен, он станет неточным; на самом деле он будет странным.
> (2) Мыслитель типа 1*, если он верит, что никогда не будет странным, становится странным.
> (3) Мыслитель типа 3, если он верит, что никогда не будет странным, окажется противоречивым.
> (4) Мыслитель типа 4, если он верит, что всегда будет непротиворечивым, станет противоречивым.

Все эти факты, в частности (4), относятся к важным математическим фактам, которые мы кратко обсудим.

Типы систем, исследованных Куртом Геделем, имели следующие особенности. Во-первых, имеется вполне определенное множество предложений, выразимых в системе. Они называются предложениями системы. Одно из этих предложений есть предложение $\perp$ (логическая ложность), и для любых предложений p и q системы предложение $p \supset q$ есть также предложение системы. Логические связки &, $\vee$, $\neg$ и $\equiv$ могут быть определены через $\supset$ и $\perp$ так, как это объяснено в главе 7.

Во-вторых, система – назовем ее S – имеет различные аксиомы и логические правила, делающие определенные предложения *доказуемыми* в системе. Таким образом, мы имеем вполне определенное подмножество множества предложений системы, а именно множество *доказуемых* предложений системы.

В-третьих, для любого предложения p в системе предложение, что p *доказуемо* в системе, само есть предложение системы (оно может быть истинно или ложно, может быть доказуемо или недоказуемо в системе). Пусть Bp будет предложением, что p доказуемо в системе. (Символ «B» для «доказуемо» был введен Геделем; он произведен от немецкого слова *beweisbar*). По счастливому совпадению, мы имеем символ «B» для использования в двух близко соотносящихся ситуациях. Когда мы говорим о *Мыслителях,* Bp означает, что Мыслитель *верит* в p; когда мы говорим о математических *системах*, Bp означает, что p *доказуемо* в системе.

Мы сейчас определим системы S типов 1, 1*, 2, 3, 4 точно таким же образом, как мы это сделали для Мыслителей: если все тавтологии доказуемы в S, и если для любых предложений p и p⊃q доказуемы в S, q также доказуемо в S, тогда система S есть система типа 1. Если также в системе S доказуемо Bp⊃Bq всякий раз, когда доказуемо p⊃q, тогда мы говорим, что это система типа 1*. Если в системе доказуемы все утверждения (Bp&B(p⊃q))⊃Bq, тогда это система типа 2. Если система также нормальна (то есть Bp доказуемо всякий раз, когда доказуемо p), тогда система типа 3. Наконец, если все предложения формы Bp⊃BBp доказуемы в S, тогда это система типа 4. Конечно, все результаты главы 11, которые мы доказали для Мыслителя, справедливы также для *систем* (где мы переинтерпретируем «B» как означающее *доказуемо*, а не *верит*). Что касается результатов из глав 10 и 12, мы нуждаемся в других условиях, к которым мы сейчас обратимся.

Геделевы системы. Гедель сделал замечательное открытие, что каждая из исследованных им систем имеет то свойство, что имеется предложение p такое, что предложение p≡¬Bp доказуемо в системе. Такие системы мы назовем *геделевыми* системами.

Предложение p≡¬Bp очень любопытно. Здесь мы имеем предложение p, эквивалентное своей собственной недоказуемости в системе! Оно может быть интерпретировано, как «Я не доказуемо в системе». Как Гедель ухитрился найти такое предложение, нам сейчас нет необходимости знать, хотя об этом будет сказано позднее.

По аналогии с системами мы могли бы определить Мыслителя как Геделева Мыслителя, если имеется по крайней мере одно предложение p такое, что он верит в предложение $p \equiv \neg Bp$. Конечно, мы тщательно изучали Геделевых Мыслителей в последней главе. (Если Мыслитель прибывает на Острове Рыцарей и Плутов, веря в правила Острова, и если Обитатель говорит ему: «Ты никогда не поверишь, что я есть Рыцарь», – тогда Мыслитель верит в предложение $k \equiv \neg Bk$, и отсюда, становится Геделевым Мыслителем.)

Давайте говорить, что система может доказать свою собственную точность, если она может доказать все предложения формы $Bp \supset p$. Мы будем говорить также, что она может доказать свою не-странность, если она может доказать все предложения формы $\neg(Bp \& B\neg Bp)$. Мы будем говорить, что система может доказать свою собственную непротиворечивость, если она может доказать предложение $\neg B\bot$.

Давайте теперь подведем итоги в терминах систем, а не Мыслителей.

Краткие итоги I*

(1) Если Геделева система типа 1 может доказать свою собственную точность, тогда она неточна, – на самом деле, она странна.

(2) Если Геделева система типа 1* может доказать свою собственную не-странность, тогда она странна.

(3) Если Геделева система типа 3 может доказать свою собственную не-странность, тогда она противоречива.

(4) Если Геделева система типа 4 может доказать свою собственную непротиворечивость, тогда она противоречива.

Пункт (4) на самом деле является важным; это обобщенная форма знаменитой Второй Теоремы о Неполноте.

Обсуждение. В статье 1931 г. Гедель имел дело с конкретной системой (системой *Principia Mathematica* Рассела и Уайтхеда) и показал, что если эта система непротиворечива, она не может доказать своей собственной непротиворечивости. Он установил, что его метод применим не только к этой конкретной системе, но и к широкому классу систем. В самом деле, метод применим ко всем Геделевым системам типа 4, как мы только что видели.

Другой важной Геделевой системой типа 4 является система, известная под названием Арифметика (более полно, Первопорядковая Арифметика Пеано). Это формализация теории обычных целых чисел 0, 1, 2, ... Поскольку Арифметика является Геделевой системой типа 4, она подвержена Теореме Геделя о Непротиворечивости; отсюда, если Арифметика непротиворечива (каковой она является, так как в ней доказуемы только истинные предложения), тогда она не может доказать своей собственной непротиворечивости.

Когда математик Андре Вейль услышал об этом, он сделал остроумный комментарий: «Бог существует, поскольку Арифметика непротиворечива; Дьявол существует, поскольку мы не можем доказать ее».

Это остроумное изречение, хотя и восхитительно, на самом деле вводит в заблуждение. Дело не в том, что мы не можем доказать непротиворечивость Арифметики, а в том, что эта Арифметика не может доказать непротиворечивости Арифметики! Мы определенно можем доказать непротиворечивость Арифметики, но наше доказательство не может быть формализовано в самой Арифметике.

На самом деле, имеется огромное число популярных изложений этой темы, которые только вводят в заблуждение по поводу Второй Теоремы о Неполноте (Теорема о Непротиворечивости). Частично это объясняется безответственностью журналистов от науки и других популяризаторов. (Я, конечно, горой за популяризацию, только чтобы она не была неточной). Один популяризатор писал: «Теорема Геделя означает, что мы никогда не узнаем, является ли Арифметика непротиворечивой». Это полный вздор. Чтобы убедиться, насколько это глупо, предположим, что Арифметика может доказать свою собственную непротиворечивость – или, если быть более реалистичным, предположим, что мы берем некоторую другую систему, которая *может* доказать свою собственную непротиворечивость. Будет ли это гарантией непротиворечивости системы? Конечно же, нет. Если бы система была противоречивой, тогда, будучи системой типа 1, она могла бы доказать все, что угодно, включая свою собственную непротиворечивость! Доверять непротиворечивости системы на том основании, что она может доказать свою собственную непротиворечивость, было бы так же глупо, как доверять правдивости человека лишь на том основании, что он говорит, что всегда говорит правду. Нет, тот факт, что Арифметика не может доказать свою собственную непротиворечивость, не бросает ни малейшей тени сомнения на непротиворечивость Арифметики.

Кстати говоря, в этой книге мы сконструируем несколько Геделевых систем типа 4 и *докажем* их непротиворечивость вне всяких сомнений. Затем мы покажем, что благодаря их собственной непротиворечивости системы неспособны доказать свою собственную непротиворечивость.

Упражнение. Рассмотрим систему S типа 4 (не необходимо Геделеву). Вспомним, что для любого предложения p, предложение Bp истинно, если и только если оно доказуемо в S.

(a) Сперва покажите, что для любого предложения p предложение $(B(p\equiv\neg Bp)\&B\neg B\bot)\supset B\bot$ истинно (для системы S). Это весьма легко.

(b) Затем покажите, что $(B(p\equiv\neg Bp)\&B\neg B\bot)\supset B\bot$ действительно доказуемо в S (это не так легко!).

ГЛАВА 14. *Еще большие затруднения с непротиворечивостью*

Некоторые предварительные проблемы

1.

Предположим, что Мыслитель верит, что он противоречив. Является ли он необходимо противоречивым? Является ли он необходимо неточным?

Решение: Предположим, что Мыслитель верит, что он противоречив. Я не вижу никакой причины, по которой он необходимо противоречив, но он должен быть неточным по следующим причинам.

Мыслитель, который верит, что он противоречив, либо прав, либо неправ в своей вере. Если его вера неверна, тогда он явно неточен (он ложно верит в противоречивость). Если его вера верна, тогда он верит в ложное предложение $\perp$. В любом случае он имеет по крайней мере одну ложную веру. Значит, он неточен.

2.

Предположим, что Мыслитель верит, что неточен. Докажите, что он прав!

Решение: Если бы он был неправ, тогда он был бы точен, что является противоречием.

3.

Предположим, что Мыслитель типа 1* верит, что он не может быть противоречивым. (Он верит в $\neg B\bot$). Будет ли он необходимо верить, что он никогда не поверит в то, что он противоречив? (То есть будет ли он необходимо верить в $\neg BB\bot$?)

Решение: Предположим, что он верит в $\neg B\bot$. Тогда он верит в логически эквивалентное предложение $B\bot\supset\bot$. Он также регулярен (поскольку он Мыслитель типа 1*), и отсюда, он будет верить в $BB\bot\supset B\bot$. Отсюда, он будет верить в логически эквивалентное предложение $\neg B\bot\supset\neg BB\bot$. Так как он также верит в $\neg B\bot$, он также верит $\neg BB\bot$.

Еще затруднения с непротиворечивостью

Мы опять на Острове Рыцарей и Плутов. Мы продолжаем предполагать, что Мыслитель принадлежит к типу 4 и что он верит в правила Острова.

4.

Предположим, что Обитатель говорит Мыслителю: «Если я – Рыцарь, тогда ты поверишь, что я – Плут». Доказать:

А) Мыслитель рано или поздно поверит, что он сам противоречив.

Б) Если правила Острова справедливы, тогда Мыслитель становится противоречивым.

Замечание: В этой главе не предполагается, что Мыслитель верит в свою непротиворечивость.

Решение: Мы знаем по Теореме 1 главы 3, что для любого предложения p, если Обитатель Острова говорит: «Если я – Рыцарь, то p», – Обитатель должен быть Рыцарем и p должно быть истинным. Теперь, любой Мыслитель – даже типа 1 – знает это точно, так же как вы и я, и поэтому если Мыслитель *верит* в правила Острова, тогда, если Обитатель говорит ему «Если я – Рыцарь, то p», Мыслитель будет верить, что Обитатель есть Рыцарь, и что p истинно. В этой конкретной проблеме Обитатель говорит: «Если я – Рыцарь, тогда ты поверишь, что я – Плут», – и поэтому Мыслитель поверит, что Обитатель есть Рыцарь, а также, что он (Мыслитель) *поверит*, что Обитатель есть Плут. И поэтому Мыслитель верит в k, а также в $B\neg k$ (k есть предложение, что Обитатель есть Рыцарь). До сих пор мы использовали только Мыслителя типа 1. Однако он типа 4, и так как он верит в k, он верит и в Bk. Таким образом, он верит в Bk и верит в $B\neg k$, но он знает, что $(Bk \& B\neg k) \supset B\bot$ (как показано в главе 11). Следовательно, он поверит в $B\bot$, то есть он поверит, что он противоречив. Он также верит, что Обитатель – Рыцарь.

До сих пор мы не использовали того факта, что справедливы правила Острова. Наш предыдущий аргумент использовал лишь тот факт, что Мыслитель *верит*, что правила справедливы. Теперь предположим, что правила на самом деле справедливы. Тогда Обитатель на самом деле Рыцарь, и Мыслитель на самом деле поверит, что Обитатель – Плут (согласно Теореме 1 Главы 3). Но так как Мыслитель верит в то, что Обитатель – Рыцарь, он получает противоречие.

5.

Предположим, что Обитатель вместо этого говорит, «Если я есть Рыцарь, тогда ты никогда не поверишь, что я есть Рыцарь». Доказать:

(a) Мыслитель никогда не будет противоречив.
(b) Правила Острова на самом деле не соблюдаются.

Замечание 1. Эта проблема имеет место, даже когда Мыслитель только типа 3.

Замечание 2. Для версии «студент и профессор теологии» последних двух проблем см. обсуждение решения Проблемы 5.

Решение: В этот раз Мыслитель верит в $k \equiv (k \supset \neg Bk)$, отсюда, он верит в k и верит в $\neg Bk$, которое является логическими следствиями $k \equiv (k \supset \neg Bk)$. Так как он верит в k, он будет верить в Bk, а, веря в $\neg Bk$, он становится противоречивым.

Если правила Острова на самом деле соблюдаются, тогда предложение $k \equiv (k \supset \neg Bk)$ – это не только то, во что верит Мыслитель, но оно является и истинным, отсюда $k \& \neg Bk$ (которое есть логическое следствие этого выражения) истинно, отсюда истинно $\neg Bk$, противореча тому факту, что Мыслитель все-таки верит, что Обитатель есть Рыцарь (и отсюда истинно скорее Bk, а не $\neg Bk$). Таким образом, правила Острова на самом деле не соблюдаются.

Обсуждение (к Проблеме 5): Давайте рассмотрим последние две проблемы в форме студента и профессора теологии. Предположим, профессор говорит: «Бог существует, но ты никогда не поверишь, что Бог существует». Если студент типа 4 и верит профессору, он должен поверить, что он противоречив. Если утверждение профессора также истинно, тогда студент на самом деле будет противоречивым.

С другой стороны, предположим, профессор говорит: «Бог существует, но ты никогда не поверишь, что Бог существует». Тогда если студент типа 4 – или даже типа 3 – и верит профессору, он будет противоречивым, и утверждение профессора также ложно.

6.

Это любопытная проблема. Мыслитель типа 4 приходит к тому, что он *верит* в Остров Рыцарей и Плутов (он верит в правила Острова), а Обитатель говорит ему следующие две вещи:

(1) Ты поверишь, что я есть Плут.
(2) Ты всегда будешь непротиворечивым.

Докажите, что Мыслитель будет противоречив и что правила Острова не соблюдаются.

Решение: Мыслитель думает так: «Предположим, он есть Плут. Тогда его второе утверждение ложно, что означает, что я буду противоречив, отсюда, я поверю чему угодно, в частности тому, что он есть Плут. Но это сделает верным его первое утверждение, делая его Рыцарем. Следовательно, противоречиво утверждать, что он – Плут, отсюда он должен быть Рыцарем. Так как он – Рыцарь, его первое утверждение истинно, отсюда, я буду верить, что он – Плут. Но сейчас я верю, что он – Рыцарь, отсюда я буду противоречив. Это доказывает, что я буду противоречив. Однако он – Рыцарь и утверждал, что я всегда буду непротиворечив. Следовательно, я никогда не буду противоречив.

В этот момент Мыслитель приходит к заключению, что он будет противоречив, и что он не будет противоречив, и поэтому он все-таки противоречив.

Так как Мыслитель становится противоречив, второе утверждение Обитателя ложно. Также, так как Мыслитель становится противоречивым, он будет верить во что угодно, включая тот факт, что Обитатель есть Плут. Это делает истинным первое утверждение Обитателя. Так как Обитатель делает одно истинное утверждение и одно ложное, правила Острова на самом деле не соблюдаются.

7.

На этот раз Обитатель делает следующие два утверждения Мыслителю типа 4:

(1) Ты никогда не поверишь, что я есть Рыцарь.
(2) Если ты когда-либо поверишь, что я есть Рыцарь, ты будешь противоречив.

Докажите, что Мыслитель становится противоречивым и что правила Острова не соблюдаются.

Решение: Мыслитель верит в следующие два предложения:

(1) $k \equiv \neg Bk$.
(2) $k \equiv (Bk \supset B\bot)$.

Поскольку он верит (1), тогда в соответствии с (b) теоремы $G^{\#}$ главы 12, Мыслитель будет верить в $Bk \supset B\bot$. И так как он верит в (2), он будет также верить в k. Поэтому, согласно (a) Теоремы $G^{\#}$, он становится противоречивым.

Далее следует, что первое утверждение Обитателя было ложным, а его второе утверждение – истинным. Следовательно, правила Острова не соблюдаются.

Застенчивые Мыслители

Давайте говорить, что Мыслитель не должен был бы верить в p, если его вера в p приведет его к противоречивости. Давайте говорить, что Мыслитель боится верить в p, если он верит в $Bp \supset B\bot$, то есть если он верит, что его вера в p приведет его к противоречивости. (Другими словами, он боится верит в p, если он верит, что он не должен верить в p).

Мы знаем, что любой Мыслитель типа 4, который верит в правила Острова, если ему Обитатель скажет, что он никогда не поверит, что тот – Рыцарь, *не должен* верить в собственную непротиворечивость. В общем, однако, нет никакой причины, по которой Мыслитель типа 4 не должен верить в собственную непротиворечивость. Но тут возникает любопытная вещь: Если по некоторой причине Мыслитель типа 4 боится верить в свою собственную непротиворечивость, уже его страх оправдывает это. Под этим я имею в виду, что любой Мыслитель типа 4, который боится верить в свою собственную непротиворечивость, на самом деле не должен верить в свою собственную непротиворечивость. Другими словами, если Мыслитель типа 4 верит, что его вера в свою собственную непротиворечивость приведет его к противоречивости, тогда это так и будет.

Как бы ни удивителен был этот факт, его не трудно доказать. Больше того, этот факт справедлив даже для нормального Мыслителя типа 1.

8.

Доказать, что если нормальный Мыслитель типа 1 боится верить в свою собственную непротиворечивость, тогда он на самом деле не должен верить в свою собственную непротиворечивость.

Замечание. В представленной выше проблеме мы видим, что для нормального Мыслителя типа 1 его вера в предложение $B\neg B\bot \supset B\bot$ является *самовыполнимой* в том смысле, что его вера в это предложение является достаточным условием для того, чтобы предложение было истинным. Тема самовыполнимых вер будет играть главную роль в следующих нескольких главах.

Решение: Мы предполагаем, что Мыслитель нормален и принадлежит к типу 1, и что он верит в $B\neg B\bot\supset B\bot$. Нам нужно показать, что, если он верит, что он непротиворечив, он становится противоречивым (и следовательно, его страх обоснован). Поэтому предположим, что он когда-либо поверит в $\neg B\bot$. Будучи нормальным, он тогда поверит в $B\neg B\bot$. Это вместе с его верой в $B\neg B\bot\supset B\bot$ приведет его к вере в $B\bot$ (потому что он типа 1). И поэтому он поверит и в $B\bot$, и в $\neg B\bot$ (он будет верить, что он противоречив и что он непротиворечив), что и делает его противоречивым.

9.

Мыслитель типа 4 есть также нормальный Мыслитель типа 1, отсюда, согласно последней проблеме, если он боится верить в свою собственную непротиворечивость, тогда он не должен верить в свою собственную непротиворечивость. Это означает, что для Мыслителя типа 4 предложение $B(B\neg B\bot\supset B\bot)\supset(B\neg B\bot\supset B\bot)$ истинно. Докажите, что любой Мыслитель типа 4 *знает*, что это предложение истинно (он знает, что если он боится верить в свою собственную непротиворечивость, тогда он не должен верить в нее).

Решение: Мыслитель рассуждает так: «Предположим, я стал бояться верить в свою собственную непротиворечивость. Это означает, что я не должен верить в свою собственную непротиворечивость, то есть я поверю в $B\neg B\bot\supset B\bot$. Так как я типа 1 (будучи типа 4), я поверю в $\neg B\bot\supset\neg B\neg B\bot$. Теперь, предполагаю, я должен также верить, что я непротиворечив, то есть предполагаю, что верю в $\neg B\bot$. Тогда, так как я верю в $B\neg B\bot \supset B\bot$, я поверю в $\neg B\neg B\bot$. Но я также верю в $B\neg B\bot$ (так как я верю в $\neg B\bot$ и я нормален), и отсюда, я противоречив. Следовательно, если я боюсь верить в свою собственную непротиворечивость, я на самом деле не могу поверить в свою собственную непротиворечивость, не впадая при этом в противоречивость. Таким образом, предложение $B(B\neg B\bot \supset B\bot) \supset \supset (B\neg B\bot \supset B\bot)$ истинно».

10.

Предположим, Мыслитель типа 4 верит, что он боится верить в свою собственную непротиворечивость. Следует ли отсюда, что он на самом деле боится верить в свою собственную непротиворечивость?

Решение: Мы только что видели, что Мыслитель верит в предложение $B(B\neg B\bot \supset B\bot) \supset (B\neg B\bot \supset B\bot)$. Следовательно, если он верит в $B(B\neg B\bot \supset B\bot)$, то он поверит в $(B\neg B\bot \supset B\bot)$. Это означает, что если он верит, что боится верить в свою собственную непротиворечивость, тогда он на самом деле будет бояться верить в свою собственную непротиворечивость.

ЧАСТЬ VI

САМОВЫПОЛНИМЫЕ ВЕРЫ И ТЕОРЕМА ЛЁБА

ГЛАВА 15. *Самовыполнимые веры*

Проблемы, обсуждаемые в этой главе, относятся к Теореме Лёба, знаменитому результату, который весьма важен для основного содержания этой книги.

Мы сейчас меняем сценарий. Мыслитель типа 4 думает о посещении Острова Рыцарей и Плутов, потому что слышал, что серные ванны и минеральные воды вылечат его ревматизм. Перед отплытием, однако, он обсуждает ситуацию со своим семейным врачом. Он спрашивает доктора, на самом ли деле лечение срабатывает. Доктор отвечает: «Лечение по большей части психологическое; вера, что оно срабатывает, является самовыполнимой. *Если ты веришь, что лечение сработает, оно сработает*».

Мыслитель доверяет доктору и поэтому приезжает на остров с априорной верой, что если он будет верить в то, что лечение срабатывает, тогда оно сработает. Он принимает лечение, которое заканчивается, но сработает, по предположению, только через несколько недель. На следующий день после лечения он начинает тревожиться по поводу результатов и думает: «Только если я могу *верить*, что лечение срабатывает, тогда оно сработает. Но откуда я знаю, поверю ли я, что оно сработает? У меня нет рациональных свидетельств, что лечение сработает, но у меня нет и свидетельств, что, если я когда-либо поверю, что оно сработает, тогда лечение сработает. Потому что я вполне могу никогда не поверить, что лечение срабатывает, и соответственно лечение не сработает!»

Мимо проходит Обитатель острова и спрашивает Мыслителя, чем тот так расстроен. Мыслитель описывает всю ситуацию, подводя итог своим размышлениям: «Если бы я когда-либо поверил, что лечение срабатывает, и оно срабатывало бы, верил бы я тогда, что лечение срабатывало?» Обитатель ответил: «*Если я когда-либо поверил, что являюсь Рыцарем, тогда ты поверил бы, что лечение срабатывает*».

С первого взгляда, это не выглядит для Мыслителя ободряющим замечанием. Он думает так: «Что это мне дает? Даже если то, что оно говорит, истинно, это все сводит проблему к тому, верю ли я, что он есть Рыцарь. Откуда мне знать, поверю ли когда-либо, что он – Рыцарь? И даже если я поверю, он все равно может быть Плутом, и его утверждение может быть ложным, и отсюда, я все еще могу не верить, что лечение срабатывает». Но вскоре Мыслитель более тщательно обдумал ситуацию и затем вздохнул с облегчением. Почему?

Ну, как мы увидим, удивительная вещь заключается в том, что Мыслитель *поверит*, что лечение сработает, и предполагая, что доктор прав, лечение сработает! Я хотел бы заметить, что правила Острова для этого аргумента не соблюдаются; достаточно того, что Мыслитель *верит* в то, что они соблюдаются.

Эта проблема близко соотносится с важной Теоремой Лёба, которую мы рассмотрим позднее. Но сперва давайте рассмотрим несколько более простую проблему, которая подходит даже ближе к исходному аргументу Лёба. Предположим, что Обитатель вместо произнесенного выше говорит: «Если ты когда-либо поверишь, что я есть Рыцарь, тогда лечение сработает».

1. (По мотивам Лёба)

Докажем, что при вышеприведенных условиях Мыслитель поверит, что лечение сработает (и отсюда, если доктор прав, лечение сработает).

Решение: Легче будет изложить решение частично словами и частично с символами. Пусть k будет предложение, что Обитатель есть Рыцарь, и пусть C будет предложение, что лечение срабатывает. Таким образом, Мыслитель верит в предложение $BC \supset C$.

Резоны Мыслителя: «Предположим, что я когда-либо поверю, что он – Рыцарь. Тогда я поверю в то, что он говорит, – я верю, что $Bk \supset C$. Также, если я когда-либо поверю, что он – Рыцарь, я поверю, что верю, что он – Рыцарь, – я поверю в Bk. И поэтому, если я когда-либо поверю, что он – Рыцарь, я поверю в оба утверждения – Bk и $Bk \supset C$, и отсюда, я поверю в C. Таким образом, если я когда-либо поверю, что он – Рыцарь, тогда я поверю, что лечение срабатывает. Но если я когда-либо поверю, что лечение срабатывает, тогда лечение срабатывает (как об этом мне сказал доктор). И поэтому, если я когда-либо поверю, что он – Рыцарь, тогда лечение сработает. Ну, а это как раз то, что и сказал. Он сказал, что если я когда-либо поверю, что он – Рыцарь, тогда лечение сработает, и он был прав! Отсюда, он – Рыцарь.

На этом этапе Мыслитель верит, что Обитатель есть Рыцарь, и поскольку Мыслитель нормален, он продолжает: «Сейчас я верю, что он – Рыцарь. Я уже доказал, что если я поверю, что он – Рыцарь, тогда лечение срабатывает, и так как я все-таки верю, что он – Рыцарь, лечение сработает».

На этом этапе Мыслитель верит, что лечение сработает. Тогда, предполагая, что доктор прав, лечение сработает.

Решение Проблемы 1 могло быть установлено более быстро, если бы мы сперва доказали следующую лемму, которая будет иметь и другие применения.

Лемма 1. Дано некоторое предложение p; предположим, что Обитатель острова говорит Мыслителю типа 4: «Если ты когда-либо поверишь, что я есть Рыцарь, тогда p истинно». Тогда Мыслитель будет верить: «Если я когда-либо поверю, что он – Рыцарь, тогда я *поверю* в p». Более обще, для двух предложений p и k, если Мыслитель типа 4 верит в утверждение $k \equiv (Bk \supset p)$, или даже в более слабое предложение $k \supset (Bk \supset p)$, тогда он верит в $Bk \supset Bp$.

Упражнение 1. Как доказывается Лемма 1? (Это та же самая проблема, которая встречается в Упражнении 3, глава 11).

Упражнение 2. Как использование Леммы 1 облегчает решение Проблемы 1?

Ответ на Упражнение 1. Давайте сперва покажем его в версии Рыцаря–Плута. Пусть k будет предложение, что Обитатель есть Рыцарь. Обитатель утверждает предложение $Bk \supset p$. Резоны Мыслителя: «Если я когда-либо поверю, что он – Рыцарь, тогда я поверю в то, что он говорит – я поверю в $Bk \supset p$. Но если я поверю в то, что он – Рыцарь, тогда я также поверю в Bk (я верю, что я верю, что он – Рыцарь). Как только я поверил в $Bk \supset p$ и поверил в Bk, я буду верить в p. И поэтому, если я когда либо поверю в то, что он – Рыцарь, тогда я поверю в p».

Конечно, более общая форма может быть доказана существенно в той же манере или же по-другому так. Предположим, что Мыслитель типа 4 верит в $k \supset (Bk \supset p)$, в которое он определенно будет верить, если поверит в более общее утверждение $k \equiv (Bk \supset p)$. Тогда в соответствии с Проблемой 2, глава 11, он поверит в $Bk \supset (BBk \supset Bp)$. Он также верит $Bk \supset BBk$. Веря в эти два предложения, он также поверит в $Bk \supset Bp$, которое является их логическим следствием. (Для любых предложений X, Y, Z, предложение $X \supset Z$ является логическим следствием $X \supset (Y \supset Z)$ и $X \supset Z$. В частности, это так, если X есть предложение Bk, Y есть предложение BBk, а Z есть предложение Bp).

Ответ на Упражнение 2. Мыслитель верит в $k \equiv (Bk \supset C)$, потому что Обитатель утверждал $Bk \supset C$. Тогда, в соответствии с Леммой 1, Мыслитель также поверит в $Bk \supset BC$. Он также верит в $BC \supset C$, и отсюда, он верит в $Bk \supset C$. Тогда он поверит в k (так как он верит в $Bk \supset C$ и $k \equiv (Bk \supset p)$), отсюда он будет верить в Bk (он нормален). Итак, он верит в Bk и верит в $Bk \supset C$, и значит, он поверит в C.

Результатом Проблемы 1 является следующая теорема.

Теорема 1 (Лёба). Для любых предложений k и C, если Мыслитель типа 4 верит в $BC \supset C$ и верит в $k \equiv (Bk \supset C)$, тогда он верит в C.

Теорема 1 дает следующий любопытный результат.

Упражнение 3. Предположим, что студент-теолог беспокоится о таких вещах, как существование Бога и своем собственном спасении. Он тогда спрашивает профессора: «Существует ли Бог?» и «Буду ли я спасен?». Профессор тогда делает следующие утверждения:

1. Если ты веришь в свое спасение, тогда ты будешь спасен.
2. Если Бог существует и ты веришь в это, тогда ты будешь спасен.
3. Если Бог не существует, тогда ты поверишь, что Бог существует.
4. Ты будешь спасен только в том случае, если Бог существует.

Доказать, принимая во внимание, что студент есть Мыслитель типа 4 и верит профессору, что: (а) студент будет верить, что будет спасен; (б) Если утверждения профессора истинны, тогда студент будет спасен.

Решение: Пусть g есть предложение, что Бог существует, а S – студент будет спасен. Тогда студент верит в следующие четыре предложения:

1. $BS \supset S$.
2. $(g \& Bg) \supset S$.
3. $\neg g \supset Bg$.
4. $S \supset g$.

Предложение $\neg g \supset Bg$ есть логическое следствие предложения (3). Это предложение, вместе с предложением (4), имеет логическим следствием предложение $(\neg Bg \vee S) \supset g$. Кроме того, $\neg Bg \vee S$ логически эквивалентно $Bg \supset S$, отсюда $(\neg Bg \vee S) \supset g$ логически эквивалентно $(Bg \supset S) \supset g$. Далее, $g \supset (Bg \supset S)$ логически эквивалентно (2), а $g \equiv (Bg \supset S)$ есть логическое следствие $(Bg \supset S) \supset g$ и $g \supset (Bg \supset S)$. Следовательно, $g \equiv (Bg \supset S)$ есть логическое следствие (1), (2) и (3). Так как студент верит в (1), (2) и (3), от также поверит в $g \equiv (Bg \supset S)$. Так как по (1) он также верит в $BS \supset S$, по Теореме 1 он поверит в S. Таким образом, BS является истинным, и если профессор прав, $BS \supset S$ истинно, отсюда истинно S, что означает, что студент спасен.

А теперь вернемся к Проблеме 1 и ее предложению: «Если ты когда-либо поверишь, что я Рыцарь, тогда лечение сработает».

2.

Предположим, что Мыслитель типа 4 опять верит в $BC \supset C$, но в этот раз Обитатель говорит: «Если ты когда-либо поверишь, что я – Рыцарь, тогда ты поверишь, что лечение срабатывает». Докажите, что Мыслитель опять поверит, что лечение срабатывает.

Решение: В этот раз Обитатель утверждает $Bk \supset BC$ (взамен $Bk \supset C$). Тогда, по Лемме 1, Мыслитель также верит в $Bk \supset BBC$ (вместо $Bk \supset BC$). Однако мыслитель верит $BC \supset C$, и так как он типа 4, он регулярен, и отсюда, он верит в $BBC \supset BC$. Веря в это и в $Bk \supset BBC$, он верит в

$Bk \supset BC$. Он также верит в $k \equiv (Bk \supset BC)$, и поэтому он поверит в k. Тогда он поверит в Bk, и поскольку он верит в $Bk \supset BC$, он поверит в BC. Но он также верит в $BC \supset C$, и отсюда, он поверит в C.

Конечно, вышеприведенный аргумент обобщается следующим образом:

Теорема 2. Даны предложения k и C. Если Мыслитель типа 4 верит в $BC \supset C$ и верит в $k \equiv (Bk \supset BC)$, тогда он верит в C.

3.

Снова Мыслитель типа 4 имеет априорную веру, что если он поверит, что лечение сработает, тогда лечение сработает. На этот раз Обитатель говорит ему: «Рано или поздно ты поверишь, что если я – Рыцарь, лечение сработает». Мы опять увидим, что Мыслитель поверит, что лечение сработает.

Более обще мы докажем следующую теорему.

Теорема 3. Для любых предложений k и C, если Мыслитель типа 4 верит в $BC \supset C$ и верит в $k \equiv B(k \supset C)$, тогда он поверит в C.

Доказательство Теоремы 3 будет облегчено следующими двумя леммами, которые представляют интерес сами по себе.

Лемма 2. Предположим, что для некоторого предложения q Обитатель говорит Мыслителю типа 4: «Ты поверишь в q». Тогда Мыслитель будет верить: «Если он – Рыцарь, тогда я поверю, что если он – Рыцарь, тогда я поверю, что он – Рыцарь». (В более абстрактной формулировке, для любых предложений k и q, если Мыслитель типа 4 верит в $k \equiv Bq$, тогда он поверит в $k \supset Bk$).

Доказательство Леммы 2. Это Проблема 9 из главы 11, которую мы уже решили, но я хочу дать ее версию в терминах Рыцарей и Плутов, которая весьма интуитивна. Обитатель говорит: «Ты поверишь в q». Мыслитель рассуждает: «Предположим, что он – Рыцарь. Тогда я *поверю* в q. Тогда я поверю, что верю в q, отсюда, я поверю в то, что но сказал, отсюда, я поверю, что он – Рыцарь. Следовательно, если он – Рыцарь, тогда я поверю в то, что он – Рыцарь».

Лемма 3. Дано некоторое предложение p, и предположим, что Обитатель говорит Мыслителю типа 4: «Ты поверишь, что если я – Рыцарь, тогда p истинно». Тогда Мыслитель поверит: «Если он – Рыцарь, тогда я *поверю* в p». (В более абстрактной формулировке, если Мыслитель типа 4 верит в $k \equiv B(k \supset p)$, тогда он поверит в $k \supset Bp$).

Доказательство Леммы 3: Мы используем Лемму 2 для того, чтобы облегчить доказательство.

Обитатель говорит: «Ты поверишь, что если я – Рыцарь, тогда p2. Пусть q будет предложением «Если я – Рыцарь, то p». Обитатель сказал Мыслителю, что тот поверит в q, и поэтому, по Лемме 2, Мыслитель будет верить, что если Обитатель ест Рыцарь, тогда он (Мыслитель) поверит, что Обитатель есть Рыцарь. И поэтому Мыслитель рассуждает так: «Предположим, что он – Рыцарь. Тогда я поверю, что он Рыцарь. Тогда я поверю в то, что он говорит – я поверю в $Bk \supset p$. Я также поверю в Bk (я поверю, что я верю, что он Рыцарь). Следовательно, если он – Рыцарь, тогда я поверю в $Bk \supset p$ и я поверю в Bk, и отсюда, я также поверю в p. И поэтому, если он Рыцарь, я поверю в p».

Доказательство Теоремы 3: Я дам доказательство в версии Рыцарь–Плут. Обитатель говорит: «Ты поверишь, что если я – Рыцарь, тогда лечение сработает». По Лемме 3, Мыслитель будет верить: «Если он – Рыцарь, тогда я поверю, что лечение сработает». Мыслитель далее продолжает: «Также, если я поверю, что лечение сработает, тогда лечение сработает. Следовательно, если он – Рыцарь, лечение сработает. Я сейчас верю, что если он – Рыцарь, тогда лечение сработает. Он сказал, что я поверю, отсюда, он – Рыцарь. И поэтому он – Рыцарь, и к тому же (как я доказал), если он – Рыцарь, тогда лечение сработает. Следовательно, лечение сработает».

На этом этапе Мыслитель поверит, что лечение сработает.

Обсуждение. Можно также легко получить Теорему 3 как следствие Теоремы 1 следующим образом. Предположим, нам даны предложения k и C такие, что Мыслитель типа 4 верит $k \equiv B(k \supset C)$, и верит в $BC \supset C$, мы должны показать, что он верит в C. Так как он верит в $k \equiv B(k \supset C)$, он должен также верить в $(k \supset C) \equiv (B(k \supset C) \supset C)$, которое есть в логическое следствие предложения $k \equiv B(k \supset C)$. Тогда он верит в предложение $k' \equiv (Bk' \supset C)$, где k' есть предложение $k \supset C$. Так как он также верит в $BC \supset C$, тогда он поверит в C по Теореме 1.

Следующее курьезное предложение иллюстрирует самореференцию, доведенную до крайности.

Упражнение 4. Предположим, что Обитатель говорит Мыслителю типа 4: «Рано или поздно ты поверишь, что если я – Рыцарь, ты поверишь, что я есть Рыцарь».

(a) Доказать, что мыслитель поверит, что Обитатель есть Рыцарь.
(b) Доказать, что если правила Острова соблюдаются, тогда Обитатель есть Рыцарь.

Решение: Это простое следствие Леммы 2.

(a) Обитатель утверждает $B\ (k \supset Bk)$. Тактим образом, имеется предложение q, а именно $k \supset Bk$ такое, что Обитатель утверждал, что Мыслитель поверит в q. Тогда, по Лемме 2, Мыслитель поверит в $k \supset Bk$. Тогда, поскольку он нормален, он поверит в $B\ (k \supset Bk)$ – он поверит в то, что сказал Обитатель. Отсюда, он поверит, что Обитатель есть Рыцарь.
(b) Так как Мыслитель верит в Bk, он определенно верит в $k \supset Bk$, отсюда, утверждение Обитателя было истинным. И поэтому Обитатель есть Рыцарь (если соблюдаются правила Острова).

Существенное математическое содержание упражнения выше состоит в том, что для любого предложения k, если Мыслитель типа 4 верит в $k \equiv B(k \supset Bk)$, тогда он также поверит в k. Если также $k \equiv B(k \supset Bk)$ истинно, истинно также и k.

Дуальные формы

Проблемы 1, 2 и 3 (более обще, Теоремы 1, 2 и 3) имеют дуальные формы, которые весьма любопытны.

1° (Дуальное к Проблеме 1)

Вновь Мыслитель типа 1 посещает Остров, уже веря, что если он поверит, что лечение срабатывает, тогда лечение сработает. Он встречает Обитателя, который говорит ему: «Лечение не сработает, и ты поверишь, что я есть Плут».

Докажите, что Мыслитель поверит, что лечение сработает.

Решение: Мы могли бы доказать это с некоторым старанием, но легче использовать преимущества Теоремы 1, которую мы уже доказали.

Обитатель утверждал $(\neg C \& B\neg k)$, и поэтому Мыслитель верит в $k \equiv (\neg C \& B\neg k)$. Но мы знаем, что $k \equiv (\neg C \& B\neg k)$ логически эквивалентно $\neg k \equiv \neg(\neg C \& B\neg k)$, которое, в свою очередь, эквивалентно $\neg k \equiv (B\neg k \supset C)$. Следовательно, Мыслитель верит в $\neg k \equiv (B\neg k \supset C)$, и поэтому он верит в предложение формы $p \equiv (Bp \supset C)$ – p тут есть предложение $\neg k$ – и поэтому, по Теореме 1, если он верит в $BC \supset C$, он поверит в C.

2° (Дуальное к Проблеме 2)

Это подобно Проблеме 1#, за исключением того, что Обитатель говорит: «Если ты поверишь, что я ест Плут, но ты никогда не поверишь, что лечение сработает». Покажите, что следует то же самое заключение (Мыслитель поверит, что лечение сработает).

3° (Дуальное к Проблеме 3)

В этот раз Обитатель говорит: «Ты никогда не поверишь, что если я есть Плут, тогда лечение сработает». (Альтернативно, он мог бы сказать: «Ты никогда не поверишь, что либо я – Рыцарь, или что лечение сработает». Покажите, что следует то же самое заключение.

Решение Проблем 2# и 3# может быть получено аналогично, через использование следствий Теорем 2 и 3 соответственно. Мы оставляем доказательство читателю.

Упражнение 5. Предположим, что Обитатель говорит Мыслителю типа 4: «Если ты когда-либо поверишь, что я Рыцарь, тогда ты будешь противоречив». Возможно ли Мыслителю верить в свою собственную непротиворечивость, не впадая при этом в противоречивость? (Подсказка: Используйте Теорему 2).

Упражнение 6. Предположим, Мыслитель типа 4 верит, что если он верит, что лечение сработает, тогда оно сработает. Предположим, мы имеем сейчас Обитателя, который говорит ему: «Если ты когда-либо поверишь, что ты поверишь, что я – Рыцарь, тогда лечение сработает».

Поверит ли Мыслитель необходимо в то, что лечение сработает?

Упражнение 7. Предположим, что вместо этого Обитатель говорит: «Ты поверишь, что если ты когда-либо поверишь, что я – Рыцарь, тогда лечение сработает».

Будет ли Мыслитель необходимо верить, что лечение сработает?

Упражнение 8. Следующий диалог происходит между студентом и профессором теологии:

Студент: Если я верю, что Бог существует, тогда поверю ли я также, что я буду спасен?

Профессор: Если это истинно, тогда Бог существует.

Студент: Если я верю, что Бог существует, буду ли я тогда спасен?

Профессор: Если Бог существует, тогда это истинно.

Докажите, что если профессор точен, и если студент верит профессору, тогда Бог должен существовать, и студент будет спасен.

Упражнение 9. Может быть доказано следующее усиление Теоремы 3. Мыслитель типа 4 прибывает на Остров для лечения и априорно имеет веру, что если он когда либо поверит, что лечение сработает, тогда оно сработает. Он спрашивает Обитателя: «Поверю ли я когда-либо, что если я – Рыцарь, тогда лечение сработает?» Обитатель отвечает: «Если это не так, лечение сработает». (Альтернативно, он мог бы ответить: «Либо это так, или же лечение сработает».) Проблема состоит в том, чтобы доказать, что Мыслитель поверит, что лечение сработает. (В более абстрактной формулировке, если Мыслитель типа 4 поверит в $k\equiv(C\vee B(k\supset C))$ и поверит в $BC\supset C$, тогда он поверит в C.) Это доказательство облегчается сперва доказательством следующих двух фактов в виде лемм:

(1) Для любых предложений p и q предположим, что Обитатель говорит Мыслителю типа 4: «Либо p истинно, или ты поверишь в q». Тогда Мыслитель поверит: «Если Обитатель есть Рыцарь, тогда я поверю, что либо p истинно, либо Обитатель есть Рыцарь».

(2) Для любых предложений p и q предположим, что Обитатель говорит Мыслителю типа 4: «Либо p истинно, либо ты поверишь, что если я ест Рыцарь, тогда q истинно». Мыслитель будет тогда верить: «Если Обитатель есть Рыцарь, тогда либо p истинно, либо я поверю, что q истинно».

Упражнение 10. Последнее упражнение имеет дуальную форму. Вновь Мыслитель типа 4 верит, что если он поверит, что лечение сработает, тогда оно сработает. Теперь он встречает Обитателя, который говорит: «Лечение не сработает, и ты никогда не поверишь, что либо я есть Рыцарь, либо лечение сработает». Докажите, что Мыслитель поверит, что лечение сработает.

ГЛАВА 16. *Алмаз Раджи*

Те из вас, кто читал восхитительный рассказ Стивенсона «Алмаз раджи», вспомнит, что в конце истории алмаз выброшен в Темзу. Однако недавние исследования показали, что алмаз впоследствии был найден Обитателем Острова Рыцарей и Плутов. Согласно одному слуху, он взял алмаз с собой в Париж и вскоре после этого умер. В соответствии с другим слухом, он отвез алмаз домой на Остров. Если верна вторая версия, тогда алмаз находится где-то на Острове.

Мыслитель типа 4 решил исследовать вторую версию в надежде найти алмаз. Он прибыл на Остров и поверил в правила Острова. Кроме того, правила Острова действительно соблюдаются. Имеется пять различных версий того, что случилось. Каждая из них представляет интерес и будет рассмотрена нами.

1. Первая версия

Согласно первой версии, когда Мыслитель достиг Острова, он встретил Обитателя, который сделал два следующих утверждения:

> (1) «Если ты когда-либо поверишь, что я есть Рыцарь, тогда ты поверишь, что алмаз находится на Острове».
> (2) «Если ты когда-либо поверишь, что я есть Рыцарь, тогда алмаз *находится* на Острове».

Если эта версия верна, находится ли алмаз на Острове?

Пусть k будет предложением, что Обитатель есть Рыцарь. Пусть D будет предложением, что алмаз находится на Острове.

Решение: Так как Обитатель сделал два утверждения, которые он сделал, тогда Мыслитель будет верить в следующие два утверждения:

(1) $k \equiv (Bk \supset BD)$;
(2) $k \equiv (Bk \supset D)$.

Мыслитель, следовательно, поверит в следующие два более слабых утверждения:

(1)' $(Bk \supset BD) \supset k$;
(2)' $k \supset (Bk \supset D)$.

Как мы увидим, факт, что Мыслитель верит даже в (1)' и (2)', достаточен для решения проблемы.

Так как он верит в (2)', по Лемме 1 из последней главы, он также поверит в $Bk \supset BD$. Веря в это и веря в (1)', он тогда поверит в k. Веря в k и веря в (2)', тогда он поверит в $Bk \supset D$. Также, поскольку он верит в k, он поверит в Bk, и отсюда, он поверит в D. Следовательно, BD истинно. Отсюда, $Bk \supset BD$ истинно, и так как (1) истинно (правила Острова действительно соблюдаются), тогда k истинно (Обитатель действительно есть Рыцарь). Тогда, поскольку (2) истинно, предложение $Bk \supset D$ истинно. Также, Bk истинно (мы видели, что Мыслитель поверит в k), и таким образом, D истинно. Следовательно, алмаз находится на Острове.

2. Вторая версия

Согласно второй, слегка отличной, версии, Обитатель взамен двух утверждений выше делает два следующих утверждения:

(a) «Если я – Плут и если ты когда-либо поверишь, что я есть Рыцарь, тогда ты поверишь, что алмаз находится на Острове».
(b) «Если я действительно Рыцарь и если ты когда-либо поверишь этому, тогда алмаз находится на Острове».

Если вторая версия верна, является ли это достаточным свидетельством для заключения, что алмаз должен быть на Острове?

Решение: Согласно этой версии, из двух утверждений Обитателя не следует, что Мыслитель поверит предложениям (1) и (2) из решения предыдущей проблемы, а следует то, что он поверит в более слабые предложения (1)' и (2)' (см. замечание выше). Но как мы видели из решения последней проблемы, этого достаточно для того, чтобы гарантировать нахождение алмаза на Острове.

Заметка: Если Обитатель говорит: «Если я есть Плут, тогда X», тогда логически следует, что если X истинно, Обитатель должен быть Рыцарем (потому что если X истинно, тогда *любое* предложение влечет X, отсюда истинно, что если Обитатель есть Плут, тогда X, но Плут не мог бы сделать такого истинного утверждения). Вот почему

Мыслитель (который верит в правила Острова) будет верить в (1)'. Что касается (2)', ясно, что если Обитатель говорит: «Если я есть Рыцарь и X», – отсюда следует, что если Обитатель есть Рыцарь, тогда X должно быть истинным.

3. Третья версия

Третья версия весьма любопытна. Согласно этой версии, Обитатель делает два следующих утверждения:

> (1) «Если ты когда-либо поверишь, что я есть Рыцарь, тогда алмаз не находится на Острове».
> (2) «Если ты когда-либо поверишь, что алмаз находится на Острове, тогда ты станешь противоречивым».

Если третья версия верна, какое заключение из нее можно сделать?

Решение:

Шаг 1. Мыслитель верит в $k \equiv (Bk \supset \neg D)$ благодаря первому утверждению Обитателя. Тогда, по Лемме 1 последней главы, Мыслитель поверит в $Bk \supset B\neg D$. И поэтому Мыслитель рассуждает так: «Если я когда-либо поверю, что он есть Рыцарь, тогда я поверю, что алмаз не находится на Острове. Если я поверю, что алмаз все-таки на Острове, тогда я буду противоречив. Следовательно, если я когда-либо поверю, что он есть Рыцарь, тогда его второе утверждение истинно. И, конечно, если его второе утверждение истинно, тогда он есть Рыцарь. Это доказывает, что если я когда-либо поверю, что он есть Рыцарь, тогда он на самом деле есть Рыцарь».

Шаг 2. Мыслитель продолжает: «Поэтому, предположим, я верю, что он есть Рыцарь. Тогда он на самом деле Рыцарь, как я только что показал, отсюда его первое утверждение $Bk \supset \neg D$ истинно. Также, если я верю, что он Рыцарь, тогда Bk истинно, и таким образом $\neg D$ истинно. Следовательно, если верю, что он Рыцарь, тогда алмаз не находится на Острове. Он сказал это только что в своем первом утверждении, и поэтому он – Рыцарь».

Шаг 3. Мыслитель продолжает: «Теперь я верю, что он Рыцарь, и я уже показал, что $Bk \supset \neg D$, отсюда $\neg D$ истинно – алмаз не находится на Острове».

Шаг 4. Мыслитель теперь верит, что алмаз не находится на Острове. Следовательно, если он когда-либо поверит, что алмаз *все-таки* на Острове, он будет противоречив. Это доказывает, что второе утверждение Обитателя было истинным, и следовательно, Обитатель на самом деле есть Рыцарь. И поэтому первое утверждение Обитателя было также истинным – $Bk \supset \neg D$ истинно. Но Bk истинно (как мы только сейчас доказали), и поэтому $\neg D$ истинно. Следовательно, алмаз не находится на Острове.

4. Четвертая версия

Согласно четвертой версии, Обитатель сделал только одно утверждение:

> (1) «Если ты когда-либо поверишь, что я есть Рыцарь, тогда ты поверишь, что алмаз находится на Острове».

Мыслитель, конечно же, остается тут ни с чем. Тогда он обсуждает эту ситуацию с Мудрецом Острова, который, как известно, является Рыцарем высшей марки. Мудрец сделал следующее утверждение:

«Если Обитатель, с которым ты говорил, есть Рыцарь и если ты когда-либо поверишь, что алмаз находится на Острове, тогда алмаз находится на Острове».

Если четвертая версия верна, какое может быть выведено заключение?

Решение:

Шаг 1: Обитатель утверждает $Bk \supset BD$, и поэтому Мыслитель верит в $k \equiv (Bk \supset BD)$. Тогда, по Лемме 1 последней главы (беря BD для p), Мыслитель будет верить в $Bk \supset BBD$, и поэтому Мыслитель рассуждает так: «Предположим, я когда-либо поверю, что Обитатель есть Рыцарь. Тогда я поверю в BD. Тогда я поверю в k и поверю в BD, отсюда я поверю в k&BD. Я также верю в утверждение Мудреца $(k\&BD) \supset D$, и поэтому, если я когда-либо поверю в k&BD, тогда я поверю в D. Следовательно, если я когда-либо поверю в k, я также поверю в D – $Bk \supset BD$ истинно. Это то, что сказал Обитатель, отсюда, он есть Рыцарь».

Шаг 2: Мыслитель продолжает: «Я теперь верю в k – Bk истинно. Также $Bk \supset BD$ истинно (как я только что доказал), отсюда BD истинно (я поверю, что алмаз находится на Острове). Таким образом, k истинно и BD истинно, поэтому k&BD истинно. Тогда, по утверждению Мудреца, D должно быть истинным – алмаз находится на Острове».

Шаг 3: Мыслитель сейчас верит в D, поэтому DB истинно. Тогда $Bk \supset BD$ определенно истинно, отсюда, Обитатель действительно Рыцарь. Таким образом, k и BD истинны оба. Отсюда, по утверждению Мудреца, D должно быть истинным – алмаз находится на Острове.

5. Пятая версия

Эта версия похожа на предыдущую, за исключением того, что Обитатель говорит:

> (1) «Ты поверишь, что если я – Рыцарь, тогда алмаз находится на Острове».

Мудрец сделал то же самое утверждение.

Предполагая, что эти пять версий равновероятны, какова вероятность, что алмаз Раджи находится на Острове?

Замечание: Математическое содержание последних двух проблем составляет усиление Теорем 2 и 3 последней главы, см. дискуссию, которая помещена за решением.

Решение: Решение этой проблемы несколько проще, чем решение Проблемы 4.

Шаг 1: Так как Обитатель сказал то, что он сказал, по Лемме 3 последней главы, Мыслитель поверит в $k \supset BD$. Отсюда, он поверит в $k \supset (k \& BD)$. Так как он также верит в утверждение Мудреца $(k \& BD) \supset D$, он поверит в $k \supset D$. Так как Обитатель сказал, что он поверит в $k \supset D$, Мыслитель поверит, что Обитатель есть Рыцарь, – он поверит в k. И так как он поверит в $k \supset D$, он поверит в D.

Шаг 2: Так как Мыслитель поверит в $k \supset D$, тогда Обитатель на самом деле Рыцарь. Отсюда, k истинно и BD также истинно (как мы только что показали). Отсюда, $k\&BD$ истинно и, так как $(k\&BD)\supset D$ истинно, D также истинно. Так что опять алмаз находится на Острове.

Сейчас мы знаем, что шансы – 80% того, что алмаз находится на Острове, так как он находится на острове в четырех из пяти случаев. Эта вероятность должна быть достаточно высокой для того, чтобы заинтересовать читателя поискать алмаз.

Обсуждение: Существенное математическое содержание Проблемы 4 состоит в том, что для любых предложений k и p, если Мыслитель типа 4 верит в $k\equiv(Bk\supset Bp)$ и верит в $(k\&Bp)\supset p$, тогда он поверит в p. Это усиление Теоремы 2 последней главы, потому что если Мыслитель типа 4 верит в $Bp\supset p$, он определенно поверит в $(k\&Bp)\supset p$, и поэтому настоящая гипотеза слабее, чем гипотеза Теоремы 2, Глава 15 (и мы вывели то же самое заключение из более слабого предположения).

Подобным же образом, существенное математическое содержание Проблемы 5 состоит в том, что если Мыслитель типа 4 верит в $k\equiv B(k\supset p)$ и $(k\&Bp)\supset p$, тогда он поверит в p. Это сильнее, чем Теорема 3 последней главы по тем же самым причинам.

Достаточно любопытно, что Теорема 1 последней главы не проявляет аналогичного усиления. Если Мыслитель типа 4 верит в $k\equiv(Bk\supset p)$ и $(k\&Bp)\supset p$, нет никаких причин заключить, что он поверит в p.

ГЛАВА 17. *Остров Лёба*

Где-то в просторах океана есть интересный остров, населенный Рыцарями и Плутами, который я буду называть Островом Лёба. Для любого человека, который посетил этот остров, и для любого предложения p существует по крайней мере один Обитатель, который говорит посетителю: «Если ты когда-либо поверишь, что я есть Рыцарь, тогда p истинно».

В проблемах этой главы, Мыслитель типа 4 посещает Остров. При этом соблюдаются правила Острова (Рыцари говорят правду, Плуты лгут, и каждый обитатель этого Острова либо Рыцарь, либо Плут), и кроме того, Мыслитель верит в правила Острова.

Проблема Генкина

Предположим, что Обитатель Острова Лёба говорит Мыслителю: «Ты поверишь, что я – Рыцарь». С первого взгляда, кажется, что нет никакого способа узнать, является Обитатель Рыцарем или Плутом. Быть может, Мыслитель поверит, что Обитатель есть Рыцарь (в этом случае Обитатель делает истинное утверждение, и следовательно, он Рыцарь). А быть может, Мыслитель никогда не поверит, что Обитатель есть Рыцарь (в этом случае Обитатель делает ложное утверждение, и соответственно он Плут). Есть ли какой-либо способ сделать выбор между двумя этими альтернативами?

Эта проблема возникает в связи со знаменитой проблемой, поставленной Леоном Хенкиным, ответ на которую был найден М. Лёбом. Удивительно то, что вполне возможно решить, кто все-таки Обитатель – Рыцарь или Плут.

1.

При заданных выше условиях, кем является Обитатель – Рыцарем или Плутом?

2.

Предположим, что Обитатель острова Лёба говорит Мыслителю типа 4: «Ты никогда не поверишь, что я – Плут». Предполагая соблюдение правил Острова (и что Мыслитель верит в них), ответить на вопрос: кем является Обитатель – Рыцарем или Плутом?

3.

Если Мыслитель типа 4 посещает Остров Лёба (и верит в правила Острова), возможно ли ему быть непротиворечивым и верить в то, что он непротиворечив?

Решения Проблем 1, 2 и 3

1. Решение довольно тонкое. Пусть P_1 будет Обитатель, который говорит: «Ты поверишь, что я – Рыцарь». Пусть k_1 будет предложение, что P_1 есть Рыцарь. Так как Мыслитель верит в правила острова, и P_1 сказал, что он верит в них, Мыслитель поверит в предложение $k_1 \equiv Bk_1$. Из этого одного лишь факта невозможно определить, является k_1 истинным или ложным; но этот остров есть Остров Лёба, отсюда, есть Обитатель P_2, который скажет Мыслителю: «Если ты когда-либо поверишь, что я – Рыцарь, тогда P_1 есть Рыцарь». (Вспомним, что для любого предложения p имеется некоторый Обитатель, который говорит Мыслителю: «Если ты когда-либо поверишь, что я – Рыцарь, тогда p».) Пусть k_2 будет предложение, что P_2 есть Рыцарь. Так как P_2 утверждает предложение $Bk_2 \supset k_1$, Мыслитель поверит в предложение $k_2 \equiv (Bk_2 \supset k_1)$. Он также поверит в $Bk_1 \equiv k_1$, отсюда, он поверит в $Bk_1 \supset k_1$ и поэтому он поверит в $Bk_1 \supset k_1$ и $k_2 \equiv (Bk_2 \supset k_1)$. Тогда, по Теореме 1 из главы 15 (подставляя k_2 вместо k и k_1 вместо C), он поверит в k_1.Так как P_1 говорит, что Мыслитель поверит в k_1, P_1 есть Рыцарь. Итак, P_1 есть Рыцарь, и Мыслитель поверит, что P_1 есть Рыцарь.

Замечание: Мы могли бы отметить, что даже без предположения, что правила острова соблюдаются, мы все еще можем заключить, что Мыслитель *поверит*, что P_1 есть рыцарь.

2. Во избежание повторения подобных аргументов давайте заметим сразу, что если Мыслитель типа 4 посещает Остров Лёба, тогда для любого предложения p, если он верит в предложение $Bp \supset p$, он поверит в p. (Основание: Так как этот остров есть Остров Лёба, некоторые Обитатели могут сказать ему: «Если ты когда-либо поверишь, что я – Рыцарь, тогда p». Отсюда, Мыслитель поверит в $k \equiv (Bk \supset p)$, где k есть предложение, что Обитатель есть Рыцарь. Тогда, так как он верит в $Bp \supset p$, он поверит в p согласно Теореме 1 из главы 15).

Теперь проблема выглядит так: «Обитатель говорит Мыслителю: «Ты никогда не поверишь, что я есть Рыцарь». Тогда Мыслитель верит в $k \equiv \neg B \supset \neg k$. Следовательно, он поверит в логически эквивалентное предложение $\neg k \equiv B\neg k$, отсюда, он поверит в $B\neg k \supset \neg k$. Следовательно, он поверит в $Bp \supset p$, при подстановке p вместо $\neg k$. Тогда, согласно Замечанию последнего раздела, он поверит в p, то есть, он поверит в $\neg k$. И поэтому Мыслитель поверит, что Обитатель есть Плут. Так как Обитатель сказал, что Мыслитель не поверит, что Обитатель есть Плут, Обитатель на самом деле есть Плут.

3. Предположим, Мыслитель типа 4 посещает Остров Лёба. Тогда, для любого предложения p, имеется некоторый Обитатель, который говорит ему: «Если ты когда-либо поверишь, что я – Рыцарь, тогда p». В частности, это верно, если мы возьмем в качестве p предложение $\perp$ (которое, как мы помним, стоит для логической ложности). Поэтому имеется Обитатель, который говорит Мыслителю: «Если ты когда-либо поверишь, что я – Рыцарь, тогда $\perp$». Таким образом, Мыслитель верит в предложение $k \equiv (Bk \supset \perp)$. Далее, $Bk \supset \perp$ логически эквивалентно $\neg Bk$. Отсюда, $k \equiv (Bk \supset \perp)$ логически эквивалентно $k \equiv \neg Bk$, и поэтому Мыслитель верит в $k \equiv \neg Bk$. Тогда, согласно Теореме 1 из главы 12, он не может верить в свою собственную непротиворечивость, не впадая при этом в противоречие.

Рефлексивность

Рефлексивные Мыслители. Мы будем говорить, что Мыслитель *рефлексивен*, если для любого предложения q имеется по крайней мене одно предложение p такое, что Мыслитель будет верить в $p \equiv (Bp \supset q)$.

Любой Мыслитель, посещающий Остров Лёба (и верящий в правила Острова), автоматически становится рефлексивным Мыслителем, так как для любого предложения q имеется по крайней мере один Обитатель, который скажет: «Если ты когда-либо поверишь, что я – Рыцарь, тогда q истинно», и поэтому Мыслитель поверит в $k \equiv (Bk \supset q)$, где k есть предложение, что Обитатель есть Рыцарь. Однако Мыслитель, который никогда не посещал Остров Лёба, мог бы быть рефлексивным Мыслителем по совершенно иным причинам (некоторые из них мы рассмотрим в главе 25).

Давайте отметим, что если *рефлексивный* Мыслитель типа 4 посещает *обычный* Остров Рыцарей и Плутов (он не должен быть Островом Лёба) и встречает Обитателя, который говорит ему: «Ты поверишь, что я – Рыцарь», тогда Мыслитель поверит, что Обитатель есть Рыцарь. (Он не нуждается во втором Обитателе, который сказал бы ему: Если я – Рыцарь, то и первый Обитатель также Рыцарь».) Также, если *рефлексивный* Мыслитель типа 4 посещает *обычный* Остров Рыцарей и Плутов, и Обитатель говорит ему: «Ты никогда не поверишь, что я – Рыцарь», – Мыслитель поверит, что Обитатель есть Плут (как в Проблеме 2).

Также никакой непротиворечивый рефлексивный Мыслитель типа 4 не может поверить, что он непротиворечив.

И еще одно: предположим, что рефлексивный Мыслитель типа 4 думает о визите на Остров Рыцарей и Плутов с серными ваннами и минеральными водами, и его семейный врач, которому он доверяет, говорит ему: «Если ты веришь, что лечение сработает, тогда оно сработает». Тогда, без дальнейших хлопот, Мыслитель поверит, что лечение сработает (ему не надо сперва ехать на остров и встречать Обитателя, который говорит ему: «Если ты поверишь, что я – Рыцарь, тогда лечение сработает»).

Все эти факты являются специальными случаями следующей теоремы (которая получается из Теоремы 1, глава 15).

Теорема А (по мотивам Лёба). Для любого предложения q, если рефлексивный Мыслитель типа 4 верит в $Bp \supset q$, он будет верить в q.

Рефлексивные системы. Давайте рассмотрим тип математических систем, описанных в главе 13. Мы вспомним, что для любой системы S, для любого предложения p, выразимого в системе, предложение Bp (p доказуемо в S) также выразимо в этой системе. (Вспомним, что «B» означает «доказуемо». Когда мы имеем только одну систему S, слово «предложение» будет пониматься как означающее «предложение, выразимое в S»).

Мы сейчас определим систему S как *рефлексивную*, если для любого предложения q имеется по крайней мере одно предложение p (выразимое в S) такое, что предложение $p \equiv (Bp \supset q)$ *доказуемо* в S. Теорема А выше явно справедлива для *систем* точно так же, как и для Мыслителей. Пусть даны некоторая рефлексивная система S и некоторое выразимое предложение q; тогда, если $Bq \supset q$ доказуемо в системе S, доказуемо и q. Это и есть Теорема Лёба.

Мы назовем S *лёбовской* системой, если для любого предложения p, если $Bp \supset p$ доказуемо в системе, доказуемо и p. Мы сейчас установили Теорему Лёба.

Теорема L (Теорема Лёба). Каждая рефлексивная система типа 4 является лёбовской, то есть для любой рефлексивной системы S типа 4, если $Bp \supset p$ доказуемо в S, доказуемо и p.

Следствие: Для любой рефлексивной системы типа 4, если $p \equiv Bp$ доказуемо, p также доказуемо.

Обсуждение: Гедель доказал свою Теорему о Неполноте для нескольких систем, включая систему Арифметики, которую мы кратко упоминали. Эти системы являются рефлексивными системами типа 4, и это позволяет проходить аргументации Геделя. Гедель сконструировал предложение g, которое утверждает свою собственную недоказуемость в системе (предложение $g \equiv \neg Bg$ доказуемо в системе).

Позднее логик Леон Генкин сконструировал предложение h такое, что $h \equiv Bh$ доказуемо в системе, и поставил проблему, есть ли какой-либо способ сказать, доказуемо ли h в системе или нет. Такое предложение может рассматриваться как утверждение: «Я доказуемо в этой системе». (Это напоминает утверждение Обитателя: «Ты *поверишь*, что я – Рыцарь»). С первого взгляда, кажется равно возможным, что h истинно и доказуемо в системе или ложно и недоказуемо в системе. Проблема оставалась нерешенной несколько лет, пока не была, наконец, разрешена Лёбом. Мы имели ответ в Следствии выше: если система рефлексивна и типа 4, тогда предложение Генкина h доказуемо в системе.

Рефлексивные и геделевы системы. Мы вспомним, что система называется *геделевской,* если имеется некоторое предложение p такое, что $p \equiv \neg Bp$ доказуемо в системе.

4.

Каждая рефлексивная система типа 1 также является геделевой системой. Почему? (Решение приведено в обсуждении Проблемы 5.)

Строгая рефлексивность

Система S строго рефлексивна, если для любого предложения q имеется предложение p такое, что $p \equiv B(p \supset q)$ доказуемо в S.

Связь между рефлексивностью и сильной рефлексивностью устанавливается следующей теоремой.

Теорема R (Теорема Рефлексивности). Она состоит из двух частей:

(a) Любая сильно рефлексивная система типа 1 рефлексивна.

(b) Любая сильно рефлексивная система типа 1* сильно рефлексивна.

5.

Доказать Теорему R.

Решения Проблем 4 и 5

4. Предположим, что S рефлексивна и типа 1. Тогда для любого предложения q имеется некоторое p такое, что $p \equiv (Bp \supset q)$ доказуемо в S. В качестве q берем предложение $\perp$, и поэтому имеется предложение такое, что $p \equiv (Bp \supset \perp)$ доказуемо в системе. Но $Bp \supset \perp$ логически эквивалентно $\neg Bp$, отсюда, $p \equiv (Bp \supset \perp)$ логически эквивалентно $p \equiv \neg Bp$, и поэтому $p \equiv \neg Bp$ доказуемо в S, и следовательно, S является геделевской системой (на самом деле, так как наша пропозициональная логика базируется на $\supset$ и $\neg$, предложение $\neg Bp$ *есть* предложение $Bp \supset \perp$, и поэтому мы не должны предполагать, что система S принадлежит к типу 1).

5. Предположим, что S типа 1. Пусть q будет любым предложением.

(a) Предположим, что S сильно рефлексивна. Тогда имеется предложение p такое, что $p\equiv B(p\supset q)$ доказуемо в S. Так как S принадлежит типу 1, $(p\supset q)\equiv\equiv(B(p\supset q)\supset q))$ доказуемо в S. (Для любых предложений X, Y и Z предложение $(X\supset Z)\equiv(Y\supset Z)$ есть логическое следствие $X\equiv Y$. Беря p в качестве X, $B(p\supset q)$ в качестве Y, и q в качестве Z, мы получаем $(p\supset q)\equiv\equiv(B(p\supset q)\supset q)$ есть логическое следствие $p\equiv B(p\supset q)$. Следовательно, имеется предложение p', а именно $p\supset q$ такое, что $p'\equiv(Bp'\supset q)$ доказуемо в S. Отсюда, S рефлексивно.

(b) Предположим также, что S регулярно (и отсюда, типа 1*) и что S рефлексивно. Тогда имеется предложение p такое, что $p\equiv(Bp\supset q)$ доказуемо в S. Так как S регулярно, $Bp\equiv B(Bp\supset q)$ доказуемо, отсюда, $(Bp\supset q)\equiv(B(Bp\supset q)\supset q)$ доказуемо. Берем в качестве $Bp\supset q$ предложение p' и видим, что $p'\equiv(Bp'\supset q)$ доказуемо. Так как имеется предложение p' такое, что $p'\equiv(Bp'\supset q)$ доказуемо в системе S, эта система сильно рефлексивна.

Замечание. Это, конечно же, следует из теоремы выше и теоремы Лёба, что любая сильно рефлексивная система типа 4 является лёбовской. Мы доказали это другим способом в Теореме 3 из главы 15.

ЧАСТЬ VII

В ГЛУБОКИХ ВОДАХ

ГЛАВА 18. *Мыслители типа G*

Скромные Мыслители

Мы называем Мыслителя *самонадеянным*, если для любого предложения p, он верит в $Bp \supset p$. Если Мыслитель верит в p, тогда нет ничего нескромного в его вере в $Bp \supset p$. (В самом деле, если Мыслитель типа 1 и верит в p, он также верит в $q \supset p$ для любого q, так как $q \supset p$ есть логическое следствие p. В частности, он будет верить в $Bp \supset p$.)

Мы будем называть Мыслителя *скромным*, если для каждого предложения p, он верит в $Bp \supset p$, только если он верит в p. (Другими словами, если он верит в $Bp \supset p$, тогда он верит в p.) По аналогии с системами, мы также могли бы назвать скромного Мыслителя *Лёбовским Мыслителем*.

Теорема Лёба устанавливает, что каждая рефлексивная система типа 4 является лёбовской. В терминологии Мыслителей, каждый рефлексивный Мыслитель типа 4 является скромным.

Многие результаты, которые могут быть доказаны для данного Мыслителя в предположении, что он рефлексивный Мыслитель типа 4, могут быть доказаны более легко при предположении, что Мыслитель есть скромный типа 4. Например, предположим, что скромный Мыслитель типа 4 (или даже скромный Мыслитель типа 1) верит, что он непротиворечив. Тогда он верит в $\neg B\bot$. Отсюда, он верит в логически эквивалентное предложение $B\bot \supset \bot$. Тогда, будучи скромным, он должен верить в $\bot$, что означает, что он противоречив! И поэтому никакой скромный Мыслитель – даже типа 1 – не может непротиворечиво верить в то, что он непротиворечив. (Это, конечно, не означает, что он необходимо противоречив. Он может оказаться непротиворечивым, но если он непротиворечив – и скромный Мыслитель типа 1 – тогда он не может верить, что он непротиворечив.)

Мыслители типа G

Давайте говорить, что Мыслитель является скромным в отношении данного предложения p, когда мы имеем тот случай, что если он верит в $Bp \supset p$, тогда он верит в p. Мыслитель является скромным, если он скромен в отношении каждого предложения p. Мы будем говорить, что Мыслитель *верит*, что он скромен в отношении p, если он верит в предложение $B(Bp \supset p) \supset Bp$ – он верит, что если он когда-либо поверит, что его вера в p влечет p, тогда он поверит в p. Мы сейчас говорим, что Мыслитель верит, что он скромен, если для любого предложения p он верит, что он скромен в отношении p, другими словами, для любого предложения p он верит в $B(Bp \supset p) \supset Bp$.

Под *Мыслителем* типа G подразумевается Мыслитель типа 4 такой, который верит, что он скромен (для любого предложения p, он верит в $B(Bp \supset p) \supset Bp$). Под *системой* типа G подразумевается система типа 4 такая, что для каждого предложения p предложение $B(Bp \supset p) \supset Bp$ доказуемо в системе.

В последние годы исследованы многие аспекты системы типа G. Джордж Булос посвятил прекрасную книгу этому предмету, которую мы весьма рекомендуем в качестве материала для дальнейшего чтения[1].

Некоторые вопросы о Мыслителях типа G возникают вполне естественно. Если Мыслитель типа G *верит*, что он скромен, является ли он необходимо скромным? Мы вскоре покажем, что ответ на этот вопрос утвердителен. В самом деле, мы увидим, что любой Мыслитель типа 1*, который верит, что он скромен, должен быть скромным. Ну, а что в отношении Мыслителя типа 4, который *является* скромным? Является ли он необходимо скромным? Мы

---

[1] *Boolos G.* The Unprovability of Consistency (Cambridge University Press, 1979).

Примечание переводчика: Через некоторое время Дж. Булос написал новую версию книги под названием The Logic of Provability (Cambridge University Press, 1993).

покажем, что ответ является утвердительным, и отсюда, что любой Мыслитель типа 4 скромен, если и только если, он верит, что он скромен, другими словами, если и только если, он типа G. Тогда мы покажем удивительный результат, который открыли независимо логики Сол Крипке, Д. де Йонг (D.H.J. de Jongh) и Джованни Самбин (Sambin), а именно, что любой Мыслитель типа 3, который верит, что он скромен, должен быть типа 4 (и отсюда, типа G).

Другой вопрос: Мы знаем, что любой рефлексивный Мыслитель типа 4 скромен (это Теорема Лёба). Является ли скромный Мыслитель типа 4 необходимо рефлексивным? То есть, если Мыслитель типа 4 скромен, необходимо ли истинно, что для любого предложения q имеется некоторое предложение p такое, что $p \equiv (Bp \supset q)$? Нетрудно доказать, что ответ на этот вопрос утвердителен, что и будет показано в следующей главе. И поэтому к концу следующей главы мы докажем, что для любого Мыслителя следующие четыре условия являются эквивалентными:

1. Он – скромный Мыслитель типа 4.
2. Он – Мыслитель типа G (он типа 4 и верит, что скромный).
3. Он – Мыслитель типа 3 и верит, что скромный.
4. Он – Мыслитель рефлексивный типа 4.

Скромность и вера в скромность

Для любого предложения p, пусть Mp есть $B(Bp \supset p) \supset \supset Bp$. Таким образом, Mp есть предложение, что Мыслитель скромен в отношении предложения p. Мы хотим показать, что любой Мыслитель типа 4, который верит, что он скромен, на самом деле скромен. Более точно, мы покажем более сильный результат, что для любого предложения p, если Мыслитель верит, что он скромен в отношении p (без необходимости веры, что он скромен в отношении любых других предложений), тогда он на самом деле скромен в отношении p (на самом деле это справедливо даже для нормальных Мыслителей типа 1). Обратное этому сильному утверждению не необходимо истинно, то есть, если Мыслитель типа 4 скромен в отношении p, тогда он не необходимо верит, что он скромен в отношении p. Однако мы покажем, что если Мыслитель типа 4 скромен в отношении предложения Mp, тогда он поверит, что он скромен в отношении p (другими словами, если он скромен в отношении Mp, тогда он поверит в Mp). Из всего этого, конечно, следует, что если Мыслитель типа 4 скромен (в отношении всех предложений), тогда он узнает, что он скромен.

1.

Почему истинно, что любой нормальный Мыслитель типа 1 (и отсюда, любой Мыслитель типа 4), который верит, что он скромен в отношении p, должен быть на самом деле скромен в отношении p?

Решение: По гипотезе, Мыслитель верит в Mp – предложение $B(Bp\supset p)\supset Bp$. Мы должны показать, что он скромен в отношении p, то есть, если он верит в $Bp\supset p$, тогда он верит в p. Поэтому мы предполагаем, что он верит в Bp, и должны показать, что он верит (или будет верить) в p.

Так как он верит в $Bp \supset p$ (по предположению), он верит в $B(Bp \supset p)$ (поскольку он нормален). Также, он верит в $B(Bp \supset p) \supset Bp$ (по гипотезе). Тогда, будучи Мыслителем типа 1, он верит, или поверит, в Bp. Он также верит в $Bp \supset p$. Отсюда, он верит, или поверит, в p.

2.

Согласно предыдущей проблеме, для любого предложения p предложение $B(Mp) \supset Mp$ *истинно* для любого Мыслителя типа 4. Доказать, что любой Мыслитель типа 4 *верит* в предложение $B(Mp) \supset Mp$. (Он знает, что если он когда-либо поверит, что скромен в отношении p, тогда он действительно скромен в отношении p.)

Доказательство: В существенных чертах, это следует из Проблемы 1 и того факта, что Мыслитель типа 4 «знает», что он типа 4, отсюда, знает, как он может мыслить. Более детально, Мыслитель рассуждает так:

Предположим, я когда-либо поверю в Mp, то есть предположим, что я когда-либо поверю в $B(Bp \supset p) \supset Bp$. Мне нужно показать, что я скромен в отношении p, то есть, если я когда-либо поверю в $Bp \supset p$, тогда я поверю в p.

И поэтому я предположу, что я поверю в $B(Bp \supset p) \supset Bp$, а также в $Bp \supset p$. Я должен тогда показать, что я поверю в p. Ну, так как я поверю в $Bp \supset p$ (мое второе предположение), я поверю в $B(Bp \supset p)$. Так как я поверю в $B(Bp \supset p) \supset Bp$ (мое первое предположение), я поверю в Bp. Раз я поверил в $Bp \supset p$ и Bp, я поверю в p».

На этом этапе Мыслитель вывел Bp из двух предположений $B(Mp)$ и $B(Bp \supset p)$, и значит, он поверит в $(B(Mp)\&\&B(Bp \supset p)) \supset Bp$ и в логически эквивалентное предложение $B(Mp) \supset (B(Bp \supset p) \supset Bp)$, которое и есть предложение $B(Mp \supset Mp$.

3.

Используя последнюю проблему, покажите, что любой Мыслитель типа 4, который скромен относительно Mp, поверит, что он скромен относительно p.

Решение: Если Мыслитель верит в $B(Mp) \supset Mp$ и скромен относительно Mp, тогда он, конечно, поверит в Mp (потому что для любого предложения q, если он верит в $Bq \supset q$ и скромен относительно q, он поверит в q – и поэтому это тот случай, если q есть предложение Mp). Сейчас Мыслитель типа 4 верит в $B(Mp) \supset Mp$ (как мы показали в прошлой Проблеме), и поэтому, если он скромен в отношении Mp, он поверит в Mp (что означает, что он поверит, что он скромен в отношении p).

Из Проблемы 1, конечно, следует, что любой Мыслитель типа 4, который верит в свою скромность, на самом деле скромен. А из Проблемы 3 следует, что если Мыслитель типа 4 скромен (в отношении любого предложения), тогда он должен верить, что он скромен (потому что для любого предложения p, он скромен в отношении Mp, и отсюда, согласно Проблеме 3, он верит, что скромен в отношении p). Мы, таким образом, установили Теорему М.

Теорема M. Мыслитель типа 4 скромен, если и только если, он верит, что скромен.

Благодаря Теореме M, Мыслитель принадлежит типу G, если и только если, он является скромным Мыслителем типа 4. В терминах систем, а не Мыслителей, мы имеем Теорему M_1.

Теорема M_1. Для системы типа 4 следующие два условия эквивалентны:

(1) Для любого предложения p, если $Bp \supset p$ доказуемо в системе, тогда доказуемо и p.

(2) Для любого предложения p предложение $B(Bp \supset p) \supset Bp$ доказуемо в системе.

Более кратко, установленный результат таков: система типа 4 является лёбовской, если и только если, она принадлежит к типу G.

Мы доказали в последней главе, что каждая рефлексивная система типа 4 является лёбовской – это и есть Теорема Лёба. Комбинируя ее с Теоремой M_1, мы имеем Теорему M_2.

Теорема M_2. Каждая рефлексивная система типа 4 есть система типа G.

Другое доказательство Теоремы M_2 будет дано в следующей главе.

Теорема Крипке, Йонга, Самбини

Мы сейчас хотим доказать, что любой Мыслитель типа 3, который верит, что он скромен, должен также верить, что он нормален, – и отсюда, должен быть типа 4, и следовательно, типа G.

На самом деле мы докажем больше. Давайте говорить, что Мыслитель нормален в отношении предложения p, если его вера в p влечет, что он также верит в Bp. Мыслитель тогда нормален, если и только если он нормален в отношении каждого предложения p. Мы будем говорить, что Мыслитель верит, что он нормален в отношении p, если он верит в предложение $Bp \supset BBp$. (Конечно, если Мыслитель нормален в отношении p, тогда предложение $Bp \supset BBp$ истинно).

Мы будем говорить, что Мыслитель верит, что он нормален, если для каждого предложения p, он верит, что нормален в отношении p. Мыслитель типа 4, тогда, есть Мыслитель типа 3, который верит, что он нормален. Теорема Крипке, Йонга и Самбина утверждает, что каждый Мыслитель типа 3, который верит, что но скромен, будет также верить что он нормален (и отсюда, будет Мыслителем типа G). Мы докажем более сильный результат, что каждый Мыслитель типа 1*, который верит, что он скромен, будет также верить, что он нормален. Но сначала мы докажем более простой результат, что каждый Мыслитель типа 1*, который скромен, также и нормален (этот результат может быть новым). Тогда мы докажем более отточенные версии этих двух результатов).

Мы вспомним из Главы 10, что используем знак C для выражения (p&Bp), и как «Мыслитель *правильно* верит в p». Предложения $Cp \supset p$, $Cp \supset Bp$, $p \supset (Bp \equiv Cp)$ являются, конечно, тавтологиями. Нам сначала нужна следующая лемма:

Лемма 1. Любой Мыслитель типа 1* верит в следующие предложения:

(a) $BCp \supset Bpp$
(b) $p \supset (BCp \supset Cp)$

4.

Почему Лемма 1 верна?

Решение. Предполагается, что Мыслитель принадлежит к типу 1*.

(a) Так как он типа 1, он верит в тавтологию $Cp \supset Bp$. Будучи регулярным, он также поверит в $BCp \supset BBp$.
(b) Так как он типа 1, он верит в тавтологию $Cp \supset p$. Будучи регулярным, он верит в $BCp \supset Bp$. Будучи типа 1, он верит в $(p\&BCp) \supset (p\&Bp)$, которое является логическим следствием последнего предложения. Таким образом, он верит в $(p\&BCp) \supset Cp$ (потому что Cp есть предложение $p\&Bp$), и, отсюда, он поверит в логически эквивалентное предложение $p \supset (BCp \supset Cp)$.

5.

Сейчас покажем, что любой скромный Мыслитель типа 1* должен быть нормальным. Более точно (и в этом состоит подсказка!), покажите, что для любого предложения p, если Мыслитель типа 1* скромен в отношении Cp, тогда он должен быть нормальным в отношении p.

Решение. Предположим, что Мыслитель типа 1* скромен в отношении Cp. Мы должны показать, что он в этом случае нормален в отношении p.

Предположим, что он верит в p. Он также верит, согласно (b) Леммы 1, в $p \supset (BCp \supset Cp)$, отсюда, он поверит в $BCp \supset Cp$. Веря в это предложение, и будучи скромным в отношении Cp, он поверит в Cp. Он также верит в тавтологию $Cp \supset Bp$, и так как он верит в Cp, он поверит в Bp.

Это доказывает, что если он верит в p, он будет верить в Bp, и поэтому будет нормален в отношении p.

6.

Сейчас покажите, что любой Мыслитель типа 1*, который верит, что он скромен, должен также верить, что он нормален. Более точно, покажите, что для любого предложения p, если Мыслитель типа 1* верит, что он скромен в отношении Cp, тогда он поверит, что он нормален в отношении p.

Решение. Предположим, что Мыслитель типа 1* верит, что он скромен в отношении Cp. По (b) Леммы 1, он верит в $p \supset (BCp \supset Cp)$. Так как он регулярен, он верит в $Bp \supset B(BCp \supset Cp)$. Но он также верит в $B(BCp \supset Cp) \supset BCp$, поскольку верит, что он скромен в отношении Cp. Веря в последние два предложения, он поверит в $Bp \supset BCp$. Он также верит в $BCp \supset BBp$, в соответствии с (a) Леммы 1, и следовательно поверит в $Bp \supset BBp$ – он также поверит, что он нормален в отношении p.

Так как любой Мыслитель типа 3 есть также Мыслимтель типа 1* (как это мы доказали в Главе 11), мы устанавливаем Теорему M_3.

Теорема M_3 (КрипкеЭ, де Йонг, Самбин). Каждый мыслитель типа 3, который верит то, что он скромен, принадлежит типу G.

Установленная а терминах систем, а не Мыслителей, Теорема M_3 утверждает, что для любой системы типа 3, если все предложения формы $B(Bp \supset p) \supset Bp$ доказуемы в системе, тогда все предложения формы $Bp \supset BBp$ доказуемы в системе, и отсюда, система должна принадлежать типу G.

Следующие два предложения предлагают любопытные альтернативные способы характеристики Мыслителей типа G.

Упражнение 1. Показать, что для любого Мыслителя типа 4, следующие два условия эквивалентны.

(a) Мыслитель принадлежит типу G.

(b) Для любых предложений p и q, Мыслитель верит в $B(q \supset p) \supset (B(Bp \supset q) \supset Bp)$.

Решение. (a) Мыслитель типа G будет рассуждать так: «Предположим, что $B(q \supset p)$ и $B(Bp \supset q)$. Это означает, что я поверю в $q \supset p$, и я поверю в $Bp \supset q$. Отсюда, я поверю в $Bp \supset p$, и так как я скромен, я поверю в p. Таким образом, $(B(q \supset p) \& B(Bp \supset q)) \supset Bp$, или что логически эквивалентно предложению, $B(q \supset p) \supset (B(Bp \supset q) \supset Bp)$».

(b) Предположим, что для любых предложений p и q, Мыслитель типа 4 верит в $B(q \supset p) \supset (B(Bp \supset q) \supset Bp)$. Оно будет истинным, если q и p одно и то же предложение, он также поверит в $B(p \supset p) \supset (B(Bp \supset p) \supset Bp)$. Он также верит в $B(p \supset p)$ – потому что он верит в тавтологию $p \supset p$, и будучи нормальным, он будет верить в $B(Bp \supset p) \supset Bp$. Следовательно, он Мыслитель типа G.

Упражнение 2. Докажите, что для любого Мыслителя типа 4, следующие два условия эквивалентны.

(a) Он принадлежит типу G.

(b) Для любых предложений p и q, если он верит в $(Bp \& Bq) \supset q$, тогда он поверит в $Bq \supset p$.

Решение. (a) Предположим, что Мыслитель типа 4 верит в $(Bp \& Bq) \supset p$. Тогда он будет размышлять так: «$(Bp \& Bq) \supset p$. Отсюда, $Bq \supset (Bp \supset p)$. Сейчас я верю в $Bq \supset (Bp \supset p)$. Теперь, предположим Bq. Тогда я поверю в Bq, и так как я верю в $Bq \supset (Bp \supset p)$, я поверю в $Bp \supset p$. И следовательно, $Bq \supset B(Bp \supset p)$. Также $B(Bp \supset p) \supset Bp$, и поэтому $Bq \supset Bp$. Таким образом, $Bq \supset (Bp \& Bq)$. И так как $(Bp \& Bq) \supset p$, получается $Bq \supset p$».

(b) Предположим, Мыслитель типа 4 таков, что для любых предложений ри q, если он верит в $(Bp \& Bq) \supset p$, тогда он верит в $Bq \supset p$». Мы покажем, что он скромен (и отсюда, принадлежит типу G).

Предположим, что он верит в $Bp \supset p$. Тогда для любого предложения q, он определенно верит в $(Bp \& Bq) \supset p$. Отсюда, для любого предложения q, он верит в $Bq \supset p$. Ну, возьмем любое предложение q такое, что он верит в q (например, q=T). Тогда, он верит в Bq, и веря в $Bq \supset p$, он поверит в p.

ГЛАВА 19. *Скромность, рефлексивность и стабильность*

Больше о Мыслителях типа G

1.

Есть очень интересный факт о непротиворечивом Мыслителе типа G или даже непротиворечивом Мыслителе типа 1*, а именно, что *не* существует предложения р такого, что Мыслитель верит, что он не верит в р! (Он не может верить в $\neg Bp$!) Почему это так?

Решение: Предположим, что скромный Мыслитель типа 1* верит в $\neg Bp$. Он также верит в тавтологию $\bot \supset p$, отсюда, он верит в $B\bot \supset Bp$ (потому что он регулярен), отсюда, он верит в логически эквивалентное предложение $\neg Bp \supset \neg B\bot$, отсюда, он верит в $\neg Bp \supset (B\bot \supset \bot)$. Так как он верит в $\neg Bp$, он верит в $B\bot \supset \bot$. Тогда, поскольку он скромен, он поверит в $\bot$, что означает, что он будет противоречив. Следовательно, если он непротиворечив (и скромный Мыслитель типа 1*), он никогда не поверит в $\neg Bp$.

2.

Отсюда следует, что если Мыслитель типа G верит, что он не верит в р, тогда он будет противоречивым. Таким образом, для Мыслителя типа G предложение $B\neg Bp\supset B\bot$ *истинно*.

Доказать, что для любого Мыслителя типа G и для любого предложения p предложение $B\neg Bp\supset B\bot$ не только истинно, но и *известно* Мыслителю как истинное.

Решение: Любой Мыслитель типа 4 (или даже нормальный Мыслитель типа 1*) последовательно поверит в следующие предложения:

1. $\bot\supset p$
2. $B\bot\supset Bp$
3. $\neg Bp\supset\neg B\bot$
4. $\neg B\bot\supset(B\bot\supset\bot)$
5. $\neg Bp\supset(B\bot\supset\bot)$
6. $B\neg Bp\supset B(B\bot\supset\bot)$

Если Мыслитель типа G, он также поверит в $B(B\bot\supset\bot)\supset$ $\supset B\bot$, отсюда, он поверит в $B\neg Bp\supset B\bot$.

3.

Эта проблема служит целям заострения внимания на предыдущей теореме. Вспомним, как мы доказывали, что каждый рефлексивный Мыслитель типа 4 есть типа G. Это делалось в два этапа: сперва мы доказали, что каждый рефлексивный Мыслитель типа 4 является Лёбовским (теорема Лёба), а затем, что каждый Лёбовский Мыслитель типа 4 является Мыслителем типа G.

Рассмотрим теперь Мыслителя типа 4, который *не* обязательно является рефлексивным. Может случиться так, что для *некоторого* предложения q имеется предложение p такое, что Мыслитель верит в $p \equiv (Bp \supset q)$, и для некоторого другого предложения q не существует такого предложения p. Итак, мы знаем следующее: если для данного q имеется такое p, что Мыслитель верит в $p \equiv (Bp \supset q)$, тогда, если Мыслитель верит в $Bq \supset q$, он также верит в q (Теорема 1 Главы 15); по этой причине предложение $B(Bq \supset q) \supset Bq$ *истинно*. Но означает ли это, что Мыслитель *знает*, что оно истинно? Ответ на этот вопрос представляет решение следующей проблемы:

Доказать, что для любых предложений p и q, если Мыслитель типа 4 верит в $p \equiv (Bp \supset q)$, тогда он *верит* в $B(Bq \supset q) \supset Bq$.

Решение: Предположим, что Мыслитель типа 4 верит в $p \equiv (Bp \supset q)$. Мы показали в Лемме 1, Глава 15, что он будет в этом случае верить в $Bp \supset Bq$. Так как он нормален, он поверит, что он верит в $p \equiv (Bp \supset q)$, и он также поверит, что верит в $Bp \supset Bq$. И поэтому он рассуждает так: «Я верю в $p \equiv (Bp \supset q)$ и верю в $Bp \supset Bq$. Теперь, предположим, что я когда-либо верил в $Bq \supset q$. Тогда, поскольку я верю в $Bp \supset Bq$, я поверю в $Bp \supset q$. И так как я верю в $p \equiv (Bp \supset q)$, я поверю в p. Тогда я поверю в Bp, и так как я верю в $Bp \supset q$, я поверю в q. Это показывает, что, если я когда-либо поверю в $Bq \supset q$, я поверю в q».

На этом этапе Мыслитель верит в $B(Bq \supset q) \supset Bq$.

Обсуждение. Конечно же, решение Проблемы 3 дает альтернативное доказательство того, что любой рефлексивный Мыслитель типа 4 должен принадлежать также типу G.

Благодаря решению Проблемы 3 можно установить, что, если даны некоторые предложения p и q, предложение $B(p\equiv(Bp\supset q))\supset(B(Bq\supset q)\supset Bq)$ *истинно* для любого Мыслителя типа 4. Можно показать, что любой Мыслитель типа 4 *знает*, что предложение выше истинно. Читатель может показать это в качестве упражнения.

Некоторые принципы неподвижной точки[2]

Мы рассмотрели два различных доказательства того, что каждый рефлексивный Мыслитель типа 4 принадлежит типу G. Мы сейчас приведем краткое доказательство того, что каждый Мыслитель типа G является рефлексивным – для каждого q имеется такое p, что он верит в $p\equiv(Bp\supset q)$. Но сначала несколько предварительных проблем.

4.

Я уже предупреждал вас об опасности веры в любое предложение формы $p\equiv\neg Bp$. Если, однако, вам случиться быть Мыслителем типа G, боюсь, что альтернативы у него попросту нет: он должен верить в него!

Итак, если дан Мыслитель типа G, найти предложение p такое, что он должен верить в $p\equiv\neg Bp$.

[2] Это специальные случаи замечательного факта, который будет обсуждаться в последней главе.

5.

Верно также, что, если дан некоторый Мыслитель типа G, имеется предложение p такое, что он верит в $p \equiv B\neg p$. Докажите это.

Решение Проблем 4 и 5: Мы сперва решим Проблему 5. Пусть p будет предложением $B\bot$. Мы утверждаем, что любой Мыслитель типа G верит в $B\bot \equiv B\neg B\bot$ (и, отсюда, верит в $p \equiv B\neg p$, где p есть предложение $B\bot$).

Мы показали в Проблеме 2, что для любого предложения p, Мыслитель типа G верит в $B\neg Bp \supset B\bot$, отсюда (беря $\bot$ для p) верит в $B\neg B\bot \supset B\bot$. Также он верит в тавтологию $\bot \supset \neg B\bot$, поэтому он верит в $B\bot \supset B\neg B\bot$. И так как он верит в $B\neg B\bot \supset B\bot$, он должен верить в $B\bot \equiv B\neg B\bot$.

А теперь решение Проблемы 4. Мы только что показали, что Мыслитель верит в $B\bot \equiv B\neg B\bot$, отсюда, он верит в $\neg B\bot \equiv \neg B\neg B\bot$. Он также верит в $p \equiv \neg Bp$, когда в качестве p выступает предложение $\neg B\bot$.

В словесном варианте, Мыслитель типа G верит в предложение, что он непротиворечив, если и только если, он не верит, что он непротиворечив. Он также верит в предложение, что он непротиворечив, если и только если, он верит, что он непротиворечив.

6.

Дано предложение q, найти предложение p (выразимое в терминах q) такое, что любой Мыслитель типа G будет верить в $p \equiv B(p \supset q)$.

Решение. Решение состоит в том, чтобы взять в качестве p предложение Bq. Давайте проверим это. Мыслитель верит в тавтологию $q\supset(Bq\supset q)$, и поскольку он регулярен, он верит в $Bq\supset B(Bq\supset q)$. Он также верит в $B(Bq\supset q)\supset Bq$ (поскольку он типа G), отсюда, он должен верить в $Bq\equiv B(Bq\supset q)$. Следовательно, он верит в $p\equiv B(p\supset q)$, где p есть предложение Bq.

7.

Сделайте то же самое с $Bp\supset q$, то есть найдите предложение p такое, что любой Мыслитель типа G поверит в $p\equiv B(p\supset q)$.

Решение: Мы показали, что Мыслитель верит в $Bq\equiv B(Bq\supset q)$. Тогда, по пропозициональной логике, он поверит в $(Bq\supset q)\equiv B(Bq\supset q)\supset q$. И поэтому, он поверит в $p\equiv(Bp\supset p)$, где p есть $Bq\supset q$. (Замечание: Мы только сейчас сдублировали доказательство Теоремы R, Глава 17. Согласно Проблеме 6, Мыслитель является строго рефлексивным, и поэтому, согласно (a) Теоремы R, он рефлексивен.)

Замечание: Результат последней проблемы состоит в том, что каждый Мыслитель типа G рефлексивен. Мы уже доказали, что каждый рефлексивный Мыслитель типа 4 принадлежит типу G, и поэтому мы сейчас имеем Теорему L*.

Теорема L*. Мыслитель типа 4 принадлежит типу G, если и только если, он рефлексивен.

Итак, каждый Мыслитель типа G является рефлексивным. Мы уже доказали, что каждый рефлексивный Мыслитель типа 4 есть типа G, и поэтому справедлива Теорема L*:

Мыслитель типа 4 есть Мыслитель типа G, если и только если, он рефлексивен.

8. Некоторые другие свойства неподвижной точки

Даны Мыслитель типа G и любые два предложения p и q, показать, что если Мыслитель верит в $p \equiv B(p \supset q)$, тогда он будет верить в $p \equiv Bq$.

Решение: Предположим, что он верит в $p \equiv B(p \supset q)$. Тогда, по Лемме 2, глава 15, он поверит в $p \supset Bq$. Он также верит в тавтологию $q \supset (p \supset q)$, и, будучи регулярным, он верит в $Bq \supset B(p \supset q)$. Также, поскольку он верит в $p \equiv B(p \supset q)$, он должен верить в $B(p \supset q) \supset p$. Так как он верит в $Bq \supset B(p \supset q)$ и $B(p \supset q) \supset p$, он верит в $Bq \supset p$. И поэтому, он верит в $Bq \supset p$ и $p \supset Bq$ (как мы показали), и отсюда, он должен верить в $p \equiv Bq$.

9.

Показать, что если Мыслитель типа G верит в $p \equiv (Bp \supset q)$, тогда он поверит в $p \equiv (Bq \supset q)$.

Решение. Предположим, что он верит в $p \equiv (Bp \supset q)$. Тогда он поверит в $p \supset (Bp \supset q)$, и, отсюда, по Лемме 1, Глава 15, он поверит в $Bp \supset Bq$. Он верит в тавтологию $q \supset (Bp \supset q)$, отсюда, он поверит в $q \supset p$ (так как он верит в $p \equiv (Bp \supset q)$). Но он регулярен, отсюда, он поверит в $Bq \supset Bp$. Поверив в это вместе с $Bp \supset Bq$, он получит $Bp \equiv Bq$. Тогда, используя пропозициональную логику, он поверит в $(Bp \supset q) \equiv (Bq \supset q)$. Тогда, так как он верит в $p \equiv (Bp \supset q)$, он поверит в $p \equiv (Bq \supset q)$.

10.

Мыслитель типа G прибывает на Остров Рыцарей и Плутов (веря в правила Острова) и спрашивает Обитателя, женат ли тот. Обитатель отвечает: «Ты поверишь, что либо я – Рыцарь, либо я женат».

Поверит ли Мыслитель с необходимостью, что Обитатель женат? Поверит ли он с необходимостью, что тот Рыцарь?

Решение: У Мыслителя нет никакой идеи, женат Обитатель или нет, но он верит, что Обитатель есть Рыцарь. Вот ход его размышления.

Пусть m будет предложение, что Обитатель женат. Обитатель утверждает $B(k \vee m)$, где k – предложение, что Обитатель есть Рыцарь. Отсюда, Мыслитель поверит в $k \equiv B(k \vee m)$. Тогда он определенно поверит в $k \supset B(k \vee m)$. Он также верит в тавтологию $k \supset (k \vee m)$, отсюда, он поверит в $Bk \supset B(k \vee m)$, и поэтому он регулярен. Он также верит в $B(k \vee m) \supset k$, так как он верит в $k \equiv B(k \supset m)$. Отсюда, он поверит в $Bk \supset k$. Тогда, будучи типа G, он поверит в k.

Стабильность

Для понимания материала следующей главы нам нужно ввести понятие *стабильности.*

Мы назовем Мыслителя *стабильным*, если для каждого предложения p, если он верит в p, тогда он на самом деле верит в p. Мы назовем Мыслителя *нестабильным*, если он не является стабильным, то есть, если имеется по крайней мере одно предложение p такое, что Мыслитель верит, что он верит в p, но на самом деле не верит в p. Конечно, каждый точный Мыслитель автоматически стабилен, но стабильность является гораздо более слабым условием, чем точность. Нестабильный Мыслитель неточен очень странным образом; в самом деле, нестабильность представляется столь же странной психологической характеристикой, как и странность.

Мы отметим, что стабильность обратна по отношению к нормальности. Если *нормальный* Мыслитель верит в p, тогда он верит в Bp, в то время как если *стабильный* Мыслитель верит в Bp, тогда он верит в p.

Мы будем использовать термины «*стабильный*» и «*нестабильный*» как для математических систем, так и для Мыслителей. Мы назовем математическую систему S стабильной, если для любого предложения p, если Bp доказуемо в S, тогда доказуемо и p. Мы будем говорить, что

Мыслитель стабилен в отношении некоторого предложения p, если предложение $BBp \supset Bp$ истинно, то есть, если его вера, что он верит в p, гарантирует, что он на самом деле верит в p. Мы скажем, что он верит в свою собственную стабильность в отношении p, если он верит в предложение $BBp \supset Bp$. Наконец, мы будем говорить, что он верит, что он стабилен, если для каждого предложения p он верит в $BBp \supset Bp$ (для каждого p, он верит, что он стабилен в отношении p).

11.

Доказать, что если скромный Мыслитель верит, что он стабилен, тогда он либо нестабилен, либо противоречив.

Решение: Предположим, что он верит в свою стабильность. Тогда для каждого предложения p, он верит в $BBp \supset Bp$, отсюда, он верит в $BB\bot \supset B\bot$. Если он скромен, тогда он поверит в $B\bot$ (потому что для любого предложения q скромный Мыслитель, который верит в $Bq \supset q$, поверит в q, и поэтому это истинно в частности, если q есть предложение $B\bot$). Так как он верит в $B\bot$, тогда, если он стабилен, он поверит в $\bot$, и таким образом будет противоречивым. Это доказывает, что, если он верит в $BB\bot \supset B\bot$, он не может быт скромным, стабильным и непротиворечивым. Следовательно, если он скромен, стабилен и непротиворечив, он никогда не сможет поверить в $BB\bot \supset B\bot$ и поэтому не сможет узнать, что он стабилен в отношении $\bot$.

Замечание. Результат выше, конечно, влечет, что никакой непротиворечивый стабильный Мыслитель типа G не сможет когда-либо узнать, что он стабилен.

ЧАСТЬ VIII

НЕВОЗМОЖНЫЕ МИРЫ

ГЛАВА 20. *Вовеки неразрешимое*

Мы обсудили Вторую Теорему о Неполноте Геделя (никакая непротиворечивая Геделевская система типа 4 не может доказать своей собственной непротиворечивости), но все еще не обсудили его Первую Теорему о Неполноте. Мы сейчас обращаемся к этой теме – сначала в контексте Мыслителя на Острове Рыцарей и Плутов.

Мы помним из последней главы, что Мыслитель называется стабильным, если для каждого предложения p, если он верит в Bp, тогда он верит в p.

Проблема неполноты

Мы будем говорить, что система вер Мыслителя *неполна*, если имеется по крайней мере одно предложение p такое, что Мыслитель никогда не поверит ни в p, ни в ¬p (он вовеки не сможет разрешить вопрос, является ли p истинным или ложным).

Следующая проблема моделирует Первую Теорему о Неполноте Геделя.

1.

Нормальный Мыслитель типа 1 прибывает на Остров Рыцарей и Плутов, веря в правила Острова. (Соблюдаются эти правила или нет, в данном случае несущественно). Он встречает Обитателя, который говорит: «Ты никогда не поверишь, что я есть Рыцарь».

Доказать, что если Мыслитель одновременно стабилен и непротиворечив, тогда его система вер неполна. Более конкретно, найти такое предложение p, чтобы соблюдались два условия:

(a) Если Мыслитель непротиворечив, тогда он никогда не поверит в p.
(b) Если Мыслитель одновременно стабилен и непротиворечив, тогда он никогда не поверит в ¬p.

Решение: Искомое предложение p – это просто предложение k, предложение, что Обитатель есть Рыцарь. Обитатель утверждает ¬Bk, отсюда, Мыслитель будет верить в k≡¬Bk.

(a) Предположим, что Мыслитель верит в k. Тогда, будучи нормальным, он поверит в Bk. Он будет также верить в ¬Bk (так как он верит в k и верит в k≡¬Bk, и принадлежит к типу 1), поэтому он будет противоречив. Следовательно, если он непротиворечив, он никогда не поверит в k.
(b) Так как Мыслитель принадлежит к типу 1 и верит в k≡¬Bk, он также верит в ¬k≡Bk. Теперь предположим, что он верит в ¬k. Тогда он поверит в Bk. Если он стабилен, он тогда поверит в k, и поэтому стано-

вится противоречивым (так как он верит в $\neg k$). Следовательно, если он одновременно стабилен и непротиворечив, он никогда не поверит в $\neg k$.

В итоге, если он одновременно стабилен и непротиворечив, он никогда не поверит, что Обитатель есть Рыцарь, и никогда не поверит, что Обитатель есть Плут.

2. Проблема, дуальная к Проблеме 1

Предположим, что Обитатель говорит: «Ты поверишь, что я есть Плут». Найти предложение p такое, что соблюдаются условия (a) и (b) Проблемы 1.

Решение: Мы видим из решения проблемы выше, что для любого предложения p (это не должно быть обязательно конкретное предложение k), если нормальный Мыслитель типа 1 верит в $p \equiv \neg Bp$, тогда, если он непротиворечив, он никогда не поверит в p, и если он также стабилен, он никогда не поверит в $\neg p$. В данной проблеме Мыслитель верит в $k \equiv B \neg k$. Следовательно, будучи Мыслителем типа 1, он верит в $\neg k \equiv \neg B \neg k$, и поэтому он верит в $p \equiv \neg Bp$, где в качестве p берется $\neg k$. Следовательно, если он непротиворечив, он никогда не поверит, что Обитатель есть Плут, и, если он также стабилен, он никогда не поверит, что Обитатель есть Рыцарь.

3. Вариант Проблемы 1

Предположим, что Обитатель говорит: «Ты поверишь, что ты никогда не поверишь, что я есть Рыцарь». Найти предложение p, удовлетворяющее условиям (a) и (b) Проблемы 1.

Решение: Искомым предложением является $\neg Bk$, как в этом легко убедиться.

Мыслитель верит в $k \equiv B\neg Bk$.

(а) Предположим, что он верит в $\neg Bk$. Тогда, будучи нормальным, он поверит в $B\neg Bk$ и будет верить в k (так как он верит в $k \equiv B\neg Bk$ и принадлежит к типу 1). Веря в k и будучи нормальным, он поверит в Bk. Таким образом, он поверит как в Bk, так и в $\neg Bk$, и поэтому он становится противоречивым. Следовательно, если он остается непротиворечивым, он никогда не поверит в Bk.

(b) Предположим, что он верит в $\neg\neg Bk$. Тогда он верит в Bk. Если он стабилен, тогда он будет верить в k. Далее, он будет верить в $B\neg Bk$ (так как он верит $k \equiv B\neg Bk$, и принадлежит типу 1). Тогда (при условии, что он стабилен) он поверит в $\neg Bk$. Таким образом, он становится противоречивым (так как он верит в $\neg\neg Bk$). Это доказывает, что если Мыслитель одновременно стабилен и непротиворечив, он никогда не поверит в $\neg\neg Bk$.

ω-непротиворечивость

Натуральное число есть по определению либо 0, либо положительное целое число 1, 2, 3, … Мы будем использовать термин «число» для «натуральных чисел». Рассмотрим свойство P (натуральных) чисел. Для любого числа n мы пишем P(n) как означающее, что n имеет свойство P. Например, если P есть свойство быть четным числом, тогда P(n) означает, что n есть четное число, и в этом случае P(0), P(2), P(4), … – истинные предложения, а P(1), P(3), P(5), …–

ложные. С другой стороны, если P есть свойство быть нечетным числом, P(0), P(2), P(4), ... будут ложными предложениями, а P(1), P(3), P(5), ...– истинными.

В стандартной символической логике символ «$\exists$» используется для фразы «существует»; символ этот называется *экзистенциальным квантором.* Для любого свойства P чисел предложение, что существует по крайней мере одно число n, имеющее свойство P, записывается в виде $\exists nP(n)$. Предположим, что у нас есть математическая система и свойство P такое, что предложение $\exists nP(n)$ доказуемо в системе, и все же для каждого конкретного n, предложеие $\neg P(n)$ доказуемо, то есть доказуемы все предложения в бесконечной последовательности $\neg P(0)$, $\neg P(1)$, $\neg P(2)$, ... Это означает, что, с одной стороны, система может доказать общее утверждение, что некоторое число имеет свойство P, и все же каждое *конкретное* число, как может быть доказано, *не* имеет этого свойства! Ясно, что неладно что-то с системой, потому что, если $\exists nP(n)$ истинно, невозможно, чтобы все предложения $\neg P(0)$, $\neg P(1)$, $\neg P(2)$, ... были также истинными. И все же такая система не является необходимо противоречивой – нельзя с необходимостью вывести формальное противоречие из всех этих предложений. Для таких систем имеется название. Они называются *ω-противоречивыми* системами. (Символ «ω» часто используется для обозначения множества натуральных чисел.)

Давайте рассмотрим следующую аналогичную ситуацию. Предположим, что некто дает вам чек, на котором написано: «Обналичивается в некотором банке». Предполагая, что в мире имеется конечное число банков, вы можете в некоторое конченое время проверить, является ли чек подлинным или подделкой; вы просто попытаетесь обналичить его в каждом банке. Если по крайней мере один банк примет чек, тогда вы знаете, что чек подлинный; если каждый банк отвергнет его, тогда у вас есть позитивное свидетельство, что чек поддельный. Но предположим, что вы живете в мире, где бесконечное число банков, каждый из который занумерован натуральным числом. Имеется Банк 0, Банк 1, Банк 2, … и т.д.

Давайте предположим, что вы бессмертны, так что для проверки чека у вас бесконечно много времени. Теперь предположим, что на самом деле ни один банк не обналичил чек. Тогда ваш чек на самом деле подделка, и все же *за конечное время вы не можете доказать это!* Вы попытаетесь сделать проверку в первой сотне миллионов банков, и все они отвергнут ваш чек. Но вы не можете считать это обстоятельство основанием для того, чтобы объявить бесчестным того, кто дал вам чек. Потому что он может всегда сказать: «Стой, не называй меня бесчестным, ты проверил еще не все банки!» И поэтому вы никогда не придете к действительному противоречию; все, что вы может получить, это ω-противоречивость. (И даже этого вы не узнаете за конечное время.)

Понятие ω-противоречивости было юмористически охарактеризовано математиком Полем Халмошем, который определил ω-противоречивую мать как такую мать, которая говорит ребенку: «Есть нечто, что ты можешь делать, но ты не можешь делать это, делать то, делать другое…» Ребенок говорит: «Но, мама, есть ли *нечто*, что я могу делать?» Мать отвечает: «О, да, но не это, не то, не другое…»

Система называется *ω-непротиворечивой*, если она не является ω-противоречивой. Таким образом, для ω-непротиворечивой системы, если $\exists nP(n)$ доказуемо, тогда имеется по крайней мере одно число n такое, что предложение $\neg P(n)$ *не* доказуемо. Противоречивая система типа 1 является также ω-противоречивой, потому что в противоречивой системе типа 1 доказуемы все предложения. Другими словами, для систем типа 1 ω-непротиворечивость автоматически влечет (обычную) непротиворечивость. Когда обсуждается ω-непротиворечивость, иногда используется термин *простая непротиворечивость* (это нужно иметь в виду во избежание путаницы). И поэтому, используя эту терминологию, ясно, что любая ω-непротиворечивая система типа 1 также просто непротиворечива.

Мы возвращаемся к исследованию Мыслителей. Во всех до сих пор рассмотренных проблемах порядок, в котором Мыслитель верил в различные предложения, не играл никакой роли. В остающейся проблеме этой главы порядок будет играть ключевую роль.

Мыслитель прибывает на Остров Рыцарей и Плутов в определенный день, который мы назовем 0-м днем. Следующий день называется 1-м днем, затем – 2-м, и т.д. Для каждого натурального числа n мы имеем n-й день; мы предполагаем, что Мыслитель бессмертен и у него впереди бесконечное число дней. Для каждого натурального числа n и некоторого предложения p предложение, в которое Мыслитель верит в n-й день, обозначается B_np. Предложение Bp, как обычно, есть предложение, что Мыслитель верит в p в тот или иной день, или, что то же самое, $\exists nB_np$ (существует некоторое n такое, что Мыслитель верит в p в n-й день). Мы называем Мыслителя ω-противоречивым, если существует по крайней мере одно предложение p такое, что Мыслитель (в то или иное время) верит в Bp, и все же для некоторого конкретного n (в то или иное время) верит в $\neg B_n p$. Мыслитель называется ω-*непротиворечивым*, если он не является ω-противоречивым.

Мы сейчас рассмотрим Мыслителя, который удовлетворяет следующим трем условиям.

Условие C_1. Он принадлежит к типу 1.

Условие C_2. Для любого натурального числа n и некоторого предложения p: (a) если Мыслитель верит в p в n-й день, он (рано или поздно) поверит в $B_n p$; (b) если он не верит в p в n-й день, он (рано или поздно) поверит в $\neg B_n p$. (Идея состоит в том, что Мыслитель помнит, в какие предложения он верил и в какие не верил в прошлые дни.)

Условие C_3. Для любого n и некоторого p Мыслитель верит в предложение $B_n p \supset Bp$ (которое, конечно, истинно).

Следующая проблема близко соотносится с оригинальной версией Первой Теоремы о Неполноте Геделя.

4. (По мотивам Геделя)

Мыслитель, удовлетворяющий трем условиям выше, прибывающий на Остров Рыцарей и Плутов, верит в правила Острова. Он встречает Обитателя, который говорит ему: «Ты никогда не поверишь, что я – Рыцарь». Доказать:

(a) Если Мыслитель (просто) непротиворечив, тогда он никогда не поверит, что Обитатель есть Рыцарь.

(b) Если Мыслитель ω-непротиворечив, тогда он никогда не поверит, что Обитатель есть Плут.

Таким образом, если Мыслитель ω-непротиворечив (и отсюда, просто непротиворечив), тогда для него догадка, является ли Обитатель Рыцарем или Плутом, останется вовеки неразрешимой.

Решение: Располагая решением Проблемы 1, самый легкий путь решения данной проблемы, состоит в том, чтобы показать, что любой Мыслитель, удовлетворяющий Условиям C_1, C_2 и C_3, а также ω-непротиворечивый, должен быть стабильным.

(a) Нужно показать, что он нормален. Предположим, что он верит в p в n-й день. Тогда, по Условию 2, он поверит в $B_n p$. Он также верит в $B_n p \supset Bp$ (по Условию 3), отсюда, будучи Мыслителем типа 1 (Условие 1), он тогда поверит в Bp. Следовательно, он нормален.

(b) Теперь предположим, что он ω-непротиворечив. Мы покажем, что он стабилен. Предположим, что он верит в Bp. Если он никогда не поверит в p, тогда для каждого числа n он не сможет поверить в p в n-й день, и отсюда, по (b) Условия 2, для каждого n он поверит в $\neg B_n p$. Но так как он верит в Bp, он будет ω-противоречивым. Следовательно, если он ω-непротиворечив и верит в Bp, он должен верить в p в некоторый день. Это доказывает, что если он ω-непротиворечив, он должен быть стабильным (предполагая, что он удовлетворяет Условиям 1, 2 и 3 – или даже просто (b) Условия 2).

Следовательно, по Проблеме 1, он останется вовеки неразрешимым.

ГЛАВА 21. *Больше неразрешимостей!*

Мыслители Россеровского типа

Гедель доказал неполноту целого семейства математических систем в предположении, что они ω-*непротиворечивы*. Дж. Баркли Россер впоследствии открыл остроумный метод демонстрации того, что эти системы неполны при более слабом условии, а именно, что они *просто* непротиворечивы. Неразрешимое предложение, сконструированное Россером, более сложное, чем у Геделя, но неразрешимость россеровского предложения устанавливается лишь в предположении простой непротиворечивости.

Давайте вернемся к Мыслителям на острове Рыцарей и Плутов, где имеет значение, в каком порядке Мыслитель верит в различные предложения. Для любых предложений p и q, мы будем говорить, что Мыслитель поверил в p *раньше*, чем он поверил в q, если имеется некоторый день, когда он верил в p и еще не верил в q. Если Мыслитель *никогда* не верил в q, но верил в p (в тот или иной день), тогда мы считаем *истинным*, что он верил в p раньше, чем верил в q. (Другими словами, ему не обязательно верить в q для того, чтобы он верил в p раньше, чем верил в q.) Пусть $Bp<Bq$ будет предложением, что Мыслитель верил в p раньше, чем в q. Если $Bp<Bq$ истинно, тогда $Bq<Bp$ явно ложно.

Мы сейчас определим Мыслителя Россеровского типа как Мыслителя типа 1, удовлетворяющего следующим условиям.

Условие R. Для любых предложений p и q, если Мыслитель верит в p в некоторый день, в который он все еще не верил в q, тогда он (рано или поздно) поверит в $Bp<Bq$ и в $\neg(Bq<Bp)$.

Идея, лежащая в основании Правила R, состоит в том, что Мыслитель имеет совершенную память в отношении того, во что он верил и во что не верил в прошлые дни. Если он верил в p перед тем, как поверил в q, тогда в первый день, когда он верил в p, он все еще не верил в q (и вероятно, никогда не поверит, а может быть, поверит когда-нибудь в будущем), и подобным образом по поводу следующего дня он вспомнит, что в первый день, когда он поверил в p, он все еще не верил в q, и поэтому он поверит в $Bp<Bq$ или в $\neg(Bpq<Bp)$.

1.

Мыслитель Россеровского типа прибывает на Остров Рыцарей и Плутов, веря в правила Острова. Он встречает Обитателя, который говорит ему: «Ты никогда не поверишь, что я Рыцарь, раньше того, как ты поверишь, что я Плут». (Символически, утверждение Обитателя таково: $\neg(Bk<B\neg k)$.)

Доказать, что если Мыслитель *просто* непротиворечив, тогда вопрос о том, является Обитатель Рыцарем или Плутом, останется для него вовеки неразрешимым.

Решение: Так как Обитатель утверждал $\neg(Bk<B\neg k)$, Мыслитель поверит в $k\equiv\neg(Bk<B\neg k)$. Предположим, что Мыслитель (просто) непротиворечив. Нам нужно показать, что он никогда не поверит в k и никогда не поверит в $\neg k$.

(a) Предположим, что он когда-либо поверит в k. Так как он непротиворечив, он никогда не поверит в $\neg k$, отсюда, он поверит в k до того, как поверит в $\neg k$. Отсюда, он поверит в $Bk<B\neg k$ (по Условию R). Но он также верит в $k\equiv\neg(Bk<B\neg k)$, отсюда, он поверит в $\neg k$, и веря в k, он станет противоречивым! Поэтому, если он непротиворечив, он никогда не поверит в k.

(b) Предположим, что он когда-либо поверит в $\neg k$. Будучи непротиворечивым, он никогда не поверит в k, отсюда, он поверит в $\neg k$ до того, как он поверит в k, отсюда, по Условию R, он поверит в $\neg(Bk<B\neg k)$. Но он также верит в $k\equiv\neg(Bk<B\neg k)$, и тогда он поверит в k и будет противоречивым. И поэтому, если он непротиворечив, он не может поверить и в $\neg k$

2.

Предположим, что Обитатель говорит: «Ты поверишь, что я – Плут до того, как ты поверишь, что я – Рыцарь». Следует ли то же самое заключение, что и в предыдущей проблеме?

Решение: Ответ утвердительный. Мы оставляем доказательство читателю.

Обсуждение: Доказуемые утверждения математических систем доказуемы на различных стадиях. Мы могли бы рассматривать математические системы как компьютерные программы для доказательств различных предложений *последовательно*. Мы говорим, что p доказуемо до q (в данной математической системе), если p доказуемо на некоторой стадии, на которой q еще не доказуемо (q могло быть доказуемо на более поздней стадии, а могло быть недоказуемо вообще). Для любых предложений p и q, выразимых в системе, предложение $Bp<Bq$ (p доказуемо до q)

также выразимо в системах того типа, который рассматривал Гедель. Россер показал, что если p доказуемо до q, тогда как предложение $Bp<Bq$, так и предложение $\neg(Bq<Bp)$ доказуемы в системе. Россер также нашел предложение p такое, что $p\equiv\neg(Bp<B\neg p)$ доказуемо в системе. (Такое предложение p соответствует Обитателю в Проблеме 1, который говорит: «Ты никогда не поверишь, что я – Рыцарь, до того как ты поверишь, что я – Плут».) Тогда, согласно аргументу в решении Проблемы 1, если p доказуемо, тогда система противоречива, и если $\neg p$ доказуемо, тогда система вновь противоречива. И поэтому, если система непротиворечива, предложение p является неразрешимым в системе.

Предложение Геделя может быть перефразировано так: «Я недоказуемо на любой стадии». Более разработанное предложение Россера может быть перефразировано так: «Я не могу быть доказано на некоторой стадии, пока ранее не доказано мое отрицание». Геделево предложение, хотя и проще, требует предположения об ω-непротиворечивости. Россеровское предложение, хотя и более сложное, срабатывает при более слабом предположении о простой непротиворечивости.

Более простая проблема неполноты

Мы обсудили два доказательства неполноты: Геделя и Россера. Есть и другое, более простое, чем эти два, доказательство, которое комбинирует метод Геделя с использованием понятия *истины* – понятия, которое впоследствии ввел логик Альфред Тарский. Для меня всегда было загадкой, почему это простое доказательство – хорошо знакомое экспертам – игнорируется в учебниках.

В следующей ниже проблеме порядок, в котором Мыслитель верит в различные предложения, не имеет никакого значения.

3.

Предположим, что мы имеем Мыслителя – назовем его Полем, – который всегда *точен* с своих верах (он никогда не верит ни в какое ложное предложение). Ему не нужно быть Мыслителем типа 1, или быть нормальным, и ему нет необходимости в действительности посещать Остров Рыцарей и Плутов. Все, что нам нужно знать о нем, так это то, что он точен.

Однажды Обитатель говорит о нем: «Поль никогда не поверит, что я – Рыцарь». Тогда логически следует, что система вер Поля должна быть неполна. Почему это так?

Решение: Если Поль когда-либо поверит, что Обитатель есть Рыцарь, это будет фальсификацией того, что сказал Обитатель, что делает Обитателя Плутом, и отсюда, делает Поля неточным в вере, что Обитатель есть Рыцарь. Но нам дано, что Поль точен, отсюда, он не сможет когда-либо поверить, что Обитатель есть Рыцарь. Отсюда, то, что сказал Обитатель, истинно, поэтому Обитатель на самом деле Рыцарь. Тогда, если Поль точен, он никогда не сможет образовать неправильной веры в то, что Обитатель есть Плут. И поэтому Поль никогда не узнает, является Обитатель Рыцарем или Плутом.

Обсуждение Решения: Математическое содержание приведенной выше проблемы таково: В системах, исследованных Геделем, мы имели не только определенные предложения, называемые *доказуемыми* предложениями, но также еще больший класс предложений, называемых *истинными* предложениями в системе. Класс истинных предложений системы отвечает истинностным таблицам для логических связок. Кроме того, для любого предложения системы предложение Bp есть *истинное* предложение системы, если и только если p есть *доказуемое* предложение системы. Гедель нашел замечательное предложение g такое, что предложение $g \equiv \neg Bg$ является *истинным* предложением системы (на самом деле, даже доказуемым в системе, но этот более сильный факт не необходим в данном аргументе). Если бы g было ложным, Bg было бы истинным, отсюда, g было бы доказуемым, отсюда, истинным, и мы получили бы противоречие. Следовательно, g истинно, отсюда, $\neg Bg$ истинно, отсюда, g не доказуемо в системе. Поэтому g истинно, но не доказуемо в системе. Так как g истинно, $\neg g$ ложно, отсюда, также недоказуемо в системе (так как все доказуемые предложения истинны). И поэтому g неразрешимо в системе.

4.

Предположим, что Обитатель говорит: «Поль однажды поверит, что я Плут». Будет ли из этого следовать, что система вер Поля неполна?

Решение: Ответ утвердительный. Мы оставляем доказательство читателю.

Более серьезное затруднение

5.

Давайте сейчас рассмотрим непротиворечивого стабильного Мыслителя типа G. Есть один очень важный вопрос, который для него должен остаться вовеки неразрешимым, а именно вопрос о его собственной непротиворечивости. Он никогда не сможет решить, непротиворечив он или нет. Почему это так?

Решение: Мы показали в Главе 18, что каждый Мыслитель типа G скромен, и мы показали, что никакой скромный Мыслитель типа 4 (или даже типа 1) не сможет поверить, что он непротиворечив. Следовательно, никакой непротиворечивый Мыслитель типа G не может узнать, противоречив он или нет.

С другой стороны, любой стабильный Мыслитель, который верит, что он противоречив, на самом деле противоречив, потому что если он верит, что он противоречив, тогда он верит в $B\bot$, и если он также стабилен, он верит в $\bot$, и отсюда, он противоречив.

Следовательно, никакой стабильный непротиворечивый мыслитель типа G не сможет никогда поверить, что он непротиворечив, или же поверить, что он противоречив. Он обречен на вечную неопределенность в этом вопросе.

Вопрос. Конечно, приведенный выше результат вполне справедлив при замене «Мыслителя» на «систему». Непротиворечивая стабильная система типа G не сможет никогда доказать ни своей собственной непротиворечивости, ни своей собственной противоречивости.

Однако при этом возникает важный вопрос: откуда мы знаем, имеются ли какие-то непротиворечивые стабильные системы типа G? Не может ли быть так, что само понятие непротиворечивой стабильной системы типа G скрывает некоторое тонкое противоречие?

Эта проблема будет полностью разрешена до того, как мы подойдем к концу этой книги.

ЧАСТЬ IX

ВОЗМОЖНЫЕ МИРЫ

ГЛАВА 22. *Это не необходимо так!*

Многое из того, что мы делали в этой книге, тесно связано с такой областью исследования, как модальная логика. Удивительный факт об этой области исследования состоит в том, что она возникла в результате чисто философских рассмотрений. Аксиоматические системы, которые возникли в ходе развития модальной логики, как оказалось, имеют самые разные интерпретации и представляют математический интерес, а также играют важную роль в теории доказательства, компьютерных исследованиях и искусственном интеллекте. В последующих главах будут рассмотрены многие интересные результаты о математических интерпретациях модальной логики.

Фундаментальной концепцией модальной логики является понятие *необходимого* истинного, а не просто истинного предложения. Мы часто говорим: «Да, то, что произошло, истинно, но на самом деле оно могло было быть и не так. *Могло быть* совсем по-другому». В других случаях мы говорим: «О, это должно быть именно так. Оно не могло быть иначе». И поэтому мы часто делаем различие между тем, что просто *оказывается* истинным, и тем, что *необходимо* истинно. Например, оказывается, что фактом является существование точно девяти планет в Солнечной системе, но вполне постижимо, что ситуация могла бы быть другой и мы имели бы больше или меньше девяти планет. С другой стороны, такое предложение, как два плюс два есть четыре, не только истинно, оно *необходимо* истинно. Ни при каких возможных обстоятельствах не могло быть истинным, что два плюс два не есть четыре.

Не вдаваясь больше в философские аспекты понятия необходимой истины, мы обратимся к некоторым логическим проблемам, иллюстрируемым так называемой семантикой Крипке, которую мы обсудим в следующей главе. Для этой цели мы установим соответствующую терминологию.

Следуя модальному логику К. Льюису, мы будем использовать букву N для *необходимо* истинного. (Более обычный символ в наше время – это □). Таким образом, для любого предложения p мы читаем Np как «p необходимо истинно». И наша нотация в предыдущих главах будет похожа на эту, за исключением того, что мы будем использовать N вместо B. Определения множества предложений типа 1, 2, 3, 4 и G будут теми же самыми, как и прежде.

1. Универсум Мыслителей

Мы сейчас рассмотрим весь универсум Мыслителей – назовем этот универсум U_1. Для любого предложения p каждый Мыслитель либо верит в p, либо не верит в p. (Неверие означает веру в ложность p.) Каждый Мыслитель не верит в $\perp$, и его веры определяются истинностными таблицами для логических связок. Например, он верит в $p\supset q$, если и только если он не верит в p или верит в q (или оба условия). Из этого следует, что каждый такой Мыслитель верит во все тавтологии. Также, если Мыслитель верит в p и верит в $p\supset q$, он должен верить в q (потому что если он не верит в q, он должен верить в p и не верить в q, отсюда, он не должен верить в $p\supset q$, вместо того чтобы верить в $p\supset q$). Из этого следует, что каждый Мыслитель есть типа 1. Нам также дано, что каждый Мыслитель знает, во что верит другой Мыслитель.

Сейчас мы встретимся с курьезной вещью. По той или иной причине, каждый Мыслитель имеет полное доверие к суждениям своих родителей, и поэтому для любого предложения p Мыслитель верит, что p *необходимо* истинно, если и только если его *родители* верят в p! Это известно как «фундаментальное правило» универсума. Это столь важно, что мы должны формально закрепить это.

Фундаментальное правило U_1: Мыслитель верит в Np, если и только если его родители верят в p.

Замечание. Ходят слухи, что сочинитель песен из нашего универсума – американский композитор – при посещении универсума U_1, услышав причину, по которой Обитатели верят в необходимость предложений, скептически покачал головой и сказал: «Это не необходимо так!» Но я не ручаюсь за истинность рассказа – это только слухи.

Предложение p называется установленным (для универсума U_1), если все Обитатели верят в него.

Ясно, что все тавтологии установлены, но множество установленных предложений выходит за пределы тавтологий. На самом деле, множество установленных предложений должно быть типа 3, то есть

(1a) Все тавтологии являются установленными предложениями.
(1b) Если p и $p \supset q$ установлены, тогда установлено и q.
(2) $(Np \& N(p \supset q)) \supset Nq$ установлено.
(3) Если установлено p, тогда установлено и Np.

Доказать, что множество установленных предложений принадлежит к типу 3.

Решение. (1) Мы знаем, что условия (1a) и (1b) истинны, так как каждый Мыслитель принадлежит к типу1.

(2) Поэтому далее мы должны продемонстрировать, что каждый Мыслитель x верит в $(Np \& N(p \supset q)) \supset Nq$, или, что то же самое, если он верит в $Np \& N(p \supset q)$, он также должен верить в Nq. Поэтому предположим, что x верит в $Np \& N(p \supset q)$. Тогда он верит в Np и $N(p \supset q)$. Так как он верит в Np, тогда его родителя верят в p. Так как он верит в $N(p \supset q)$, его родители верят в $p \supset q$. Следовательно, его родителя верят в p и в $p \supset q$, и, будучи типа 1, они верят в q. Так как его родители верят в q, x верит в Nq.

Таким образом, мы доказали, что любой Мыслитель x из U_2 верит в $(Np \& N(p \supset q)) \supset Nq$, и поэтому это предложение установлено.

(3) Наконец, мы должны показать, что если p установлено, установлено и Np. (Это не означает, что каждый Мыслитель, который верит в p, будет также верить в Np, но только то, что если *все* Мыслители верят в p, тогда *все* они верят в Np). Это абсолютно ясно. Предположим, что все Мыслители верят в p. Возьмем любого Мыслителя x. Тогда его родители верят в p (потому что все Мыслители верят в это), отсюда, x верит в Np.

2. Второй универсум

Сейчас мы посещаем другой универсум, который назовем U_2. Условия, определяющие этот универсум, те же, что и для универсума U_1, с одним важным отличием. В этом универсуме Мыслитель верит, что p необходимо истинно, если и только если все его *предки* верят в p. (Для упрощения дела, мы предположим, что все Обитатели универсума бессмертны, так что все предки некоторого человека живы.)

Давайте запишем этот фундаментальный факт.

Факт 2. В универсуме U_2 Мыслитель x верит в Np, если и только если все предки x верят в p.

Сейчас вещи становятся более интересными.

Доказать, что множество установленных предложений универсума U_2 должно быть типа 4.

Решение: Важность изменения «родителей» на «предков» состоит в следующем: Если y есть родитель x и z – родитель y, мы едва ли можем заключить, что z есть родитель x. Но если y есть предок x и z есть предок y, тогда z есть предок x. (В математической терминологии, отношение «быть предком» транзитивно.)

Доказательство того, что множество установленных предложений в универсуме U_2 есть типа 3, то же самое, что и доказательство для универсума U_1 (просто заменить «родитель» на «предок»). Но сейчас мы можем доказать дополнительный факт, что каждый Мыслитель этого универсума верит в $Np \supset NNp$ и, отсюда, что множество установленных утверждений типа 4.

Предположим, что x верит в Np. Мы должны показать, что он должен верить также в в NNp. Пусть x' будет любым предком x, и пусть x'' будет любым предком x'. Тогда x'' будет предком x. Так как x верит в Np, а x'' есть предок x, x'' должен верить в p. Таким образом, каждый предок x'' предка x' верит в p, отсюда, x верит в p, отсюда, x верит в Np. Это показывает, что каждый предок x' индивида x верит в Np, и поэтому x должен верить в NNp.

3. Третий универсум

До сих пор мы оставляли открытым вопрос, имеет ли наш универсум начало во времени. Ну, мы рассмотрим третий универсум U_3, удовлетворяющий всем условиям U_2, плюс условие, что он имеет начало во времени. Это означает, что для каждого индивида x, если мы возьмем предка x' индивида x, а затем предка x'' индивида x' и будем продолжать в том же духе, рано или поздно мы должны прийти к предку, который не имеет предков, и отсюда, родителей. (Ответить на вопрос, как возникли эти индивиды без родителей, мы оставляем за пределами этой книги. Интересующийся читатель должен обратиться к книгам по эволюции или же креационизму в зависимости от его научных или теологических интересов.)

Как читатель может догадаться, мы имеем целью показать, что установленные предложения этого универсума образуют класс типа G. Но перед этим мы должны прояснить читателю, не знакомому с этими материями, ситуацию с логикой *все* и *некоторый*.

Предположим, что кто-то говорит об определенном клубе: «Все французы в этом клубе носят береты», – и оказывается, что в этом клубе вообще нет французов. Должно ли это предложение считаться истинным, ложным или же эти понятия к нему неприложимы? Не знакомые с формальной логикой могут иметь различные мнения по этому вопросу, но соглашение по этому поводу, принятое в логике, математике и естественных науках, говорит, что любое утверждение формы «Все A есть B» считается ложным, *только* если имеется по крайней мере один A, который не есть B. И поэтому единственный вариант, когда утверждение «Все французы в этом клубе носят береты» может быть ложным, это когда имеется, по крайней мере, один француз в клубе, который не носит берета. Если случится так, что в клубе нет французов, тогда определенно в клубе нет французов, которые не носят беретов, и следовательно, *истинно*, что «Все французы в этом клубе носят береты». Эту конвенцию мы примем в дальнейшем.

В применении к универсуму U_3, если x есть индивид, не имеющий предков, тогда все, что можно сказать о *всех* его предках, истинно (потому что он не имеет их!). В частности, если дано предложение p, мы считаем истинным, что все предки x верят в p, и поэтому, если x не имеет предков, тогда x верит в Np. (*Единственный* вариант, когда индивид x не верит в Np, это тогда, когда он имеет по крайней мере одного предка, который не верит в p, что невозможно для индивида, который вообще не имеет предков.) Давайте рассматривать это как Факт 1.

Факт 1. Если x не имеет предков, тогда для любого предложения p, x верит в Np.

Мы стремимся показать, что множество установленных предложений U_3 есть множество типа G. Оно определенно типа 4 (по Проблеме 2, так как все условия для U_2 справедливы для U_3). Тогда остается показать, что для любого предложения p все Обитатели U_3 верят в предложение $N(Np \supset p) \supset Np$. Доказательство этого очень изящное, и ключевая идея содержится в следующей лемме.

Лемма 1. Если x не верит в Np, тогда x должен иметь предка y, который не верит в p и верит в Np.

Прежде всего, докажите эту Лемму. Затем покажите, что множество установленных предложений U_3 есть множество типа G.

Решение: Чтобы доказать Лемму, предположим, что x не верит в Np. Тогда он должен иметь, по крайней мере, одного предка x', который не верит в p (потому что если все его предки верят в p, он должен верить в Np, чего он не делает). Теперь, если x' верит в Np, тогда все в порядке – надо просто взять y вместо x'. Но если x' не верит в Np, x' должен иметь по крайней мере одного предка x'', который не верит в p. Если x'' верит в Np, все в порядке (берем y вместо x”). Но если это не так, тогда мы берем некоторого предка x’’’ индивида x”, который не верит в p, и продолжаем в такой манере до тех пор, пока не достигнем некоторого предка y индивида x, который не верит в p и который либо не имеет предков вообще (в этом случае y верит в Np согласно Факту 1), или же имеет предков, все из которых верят в p, и отсюда, y должен верить в Np. (Причина, по которой мы, в конце концов, должны достигнуть такого предка y, состоит в том, что универсум U_3 имеет определенное начало во времени – факт, который не дан для универсума U_2.)

Сейчас мы можем доказать, что все Мыслители этого универсума должны верить в $N(Np \supset p) \supset Np$ (и отсюда, что множество установленных предложений есть множество типа G). Чтобы доказать это, достаточно показать, что каждый Мыслитель, который верит в $N(Np \supset p)$, будет верить в Np; или же, что то же самое, любой Мыслитель, который не в Np, не будет также верить в $N(Np \supset p)$.

И поэтому предположим, что x не верит в Np. Тогда по Лемме 1, x имеет предка y, который не верит в p и верит в Np. Тогда он верит, что Np истинно, а p ложно, и поэтому он должен не верить в $Np \supset p$. Следовательно, x имеет предка y, который не верит в $Np \supset p$, отсюда, не все предки индивида x верят в $Np \supset p$, и поэтому x должен *не верить* в $N(Np \supset p)$.

Это доказывает, что если x не верит в Np, он не верит в $N(Np \supset p)$, отсюда, если x верит в $N(Np \supset p)$, он должен верить в Np, и следовательно, x должен верить в $N(Np \supset p) \supset$ $\supset Np$.

Как все это связано с семантикой Крипке, будет объяснено в следующей главе.

ГЛАВА 23. *Возможные миры*

Предмет модальной логики весьма стар – он восходит, по крайней мере, к Аристотелю. Главными понятиями модальной логики являются понятия «*необходимо*» истинного предложения и «*возможно*» истинного предложения. Каждое из этих понятий может быть определено в терминах другого понятия. Если мы начинаем с понятия необходимой истины, мы можем определить *возможно* истинное предложение как такое, которое не является необходимо ложным. Альтернативно, мы можем начать с возможно истинного предложения и определить необходимо истинное предложение как такое, которое не является возможно ложным.

Модальная логика привлекала средневековых философов и теологов, а позднее стала фундаментальной в философии Лейбница. Именно Лейбниц вдохновил современного философа Сола Крипке, изобретшего область, известную под названием *семантика возможных миров*, также называемую *семантикой Крипке* (с которой мы работали в предыдущей главе, используя другую терминологию).

У Лейбница была идея, что мы обитаем в месте, называемом действительным миром, который представляет собой только один из *возможных* миров. Согласно теологии Лейбница, Бог сначала обозрел *все* возможные миры и затем актуализировал один из них, который посчитал лучшим, а именно – этот мир. Отсюда знаменитый афоризм «Этот мир является наилучшим из всех возможных».

(В *Кандиде* Вольтер обсмеял эту идею. После описания каждой мыслимой катастрофы Вольтер всегда добавлял: «в лучшем из всех возможных миров»).

Согласно Лейбницу, предложение p мыслилось истинным *для* или *в* данном мире x (в действительном или же возможном мире), если оно правильно описывает этот мир, и ложным, если он не делает этого. Если предложение p объявляется истинным без оговорки, это означает, что оно истинно для этого (действительного) мира. Лейбниц называл p *необходимо* истинным, если оно истинно для всех возможных миров, и *возможно* истинным, если оно истинно для по крайней мере одного возможного мира. Такова была – в кратце – философия возможных миров Лейбница. (Интересно, если бы был актуализирован некоторый другой возможный мир, имел бы Лейбниц ту же самую философию?)

До 1910 г. трактовка модальной логики была лишена точности, присущей другим отраслям логики. Даже Аристотель, теория силлогизмов которого совершенно ясна, не дал четкого объяснения модальной логики. Американский философ К. Льюис в серии статей, опубликованных между 1910 и 1920 гг., описал последовательность аксиоматических систем и исследовал, какие предложения доказуемы в каждой из них. В каждой из этих систем все тавтологии являются аксиомами (по крайней мере, доказуемы из них), и для любых предложений p и q Льюис принял правило, что если p и $p \supset q$ доказуемы, тогда доказуемо и q. Поэтому все системы Льюиса, по крайней мере, типа 1. Далее, Льюис предположил, что если p и $p \supset q$ необходимо истинны, тогда необходимо истинно и q. Отсюда, все предложения формы $(Np\&N(p\supset q))\supset Nq$ (или же, альтернативно, предложения формы $N(p\supset q)\supset(Np\supset Nq))$ брались в качестве аксиом. Далее, вполне разумно было предположить, что все, что может быть *доказано* на основании необходимо истинных аксиом, должно быть необходимо истинно, и сегодня в большинстве модальных систем (так называемых нормальных систем) в качестве правила вывода принимается, что, если предложение X *доказано*, мы вправе заключить, что NX.

Это не означает, что $X\supset NX$ необходимо истинно, но значит лишь, что если X *доказано* – на основании аксиом, которые сами должны быть необходимо истинными, – тогда мы обоснованы в утверждении NX. Описанная выше система принадлежит к типу 3 и имеет в наши дни стандартное название. Она называется *модальной системой* K. Эта система служит основанием широкого класса модальных систем. Ну, а что с предложением $NX\supset NNX$? (Если X необходимо истинно, то необходимо, что оно необходимо истинно?) Ну, предложения этой формы берутся в качестве аксиом в некоторых модальных системах, а в некоторых таковыми не являются. Модальная система, в которых к аксиомам системы K присоединяется аксиома $NX\supset NNX$, является очень важной, и в наши дни носит название модальной системы K_4. Это система явно типа 4.

Модальная система G, которая получается добавлением к K_4 всех предложений формы $(Np\&N(p\supset q))\supset Nq$ в качестве аксиом, возникла только в середине 1970-х и не из-за философского интереса к понятию логической необходимости, а из-за Теоремы Геделя и Теоремы Леба. Более подробно об этих вещах рассказано в следующей главе.

Модели Крипке

В конце 1950-х гг. Сол Крипке опубликовал свою знаменитую статью «*Теорема полноты в модальной логике*», которая открыла новую эру в модальной логике. Впервые для модальных систем была предложена точная теория моделей, которая представляла не только *математический* интерес, но которая привела к возникновению целой отрасли философии, известной сегодня под названием семантики возможных миров.

Крипке первым поднял вопрос о системах Лейбница, который Лейбниц сам не рассматривал. Согласно Лейбницу, мы обитаем в действительном мире. А являются ли так называемые возможные миры существующими мирами или же они возможны только *относительно действительного мира*? Другими словами, с точки зрения другого мира, является ли класс возможных миров отличным от класса миров, которые возможны относительно действительного мира? Предположим, что мы даем описание некоторого мира x и рассмотрим предложение «x есть возможный мир». Является ли истинность или ложность *этого* предложения чем-то абсолютным или же, может быть, что одно и то же предложение истинно в некотором мире y, но ложно в некотором другом мире z? Важным вопросом тут является вопрос о транзитивности: если мир y возможен относительно мира x, и если мир z возможен относительно мира y, следует ли отсюда, что мир z возможен относительно мира x? Ответ на этот вопрос определяет, какая из систем модальной логики является подходящей.

Следуя Крипке, мы будем говорить, что мир y *достижим от* мира x, если y возможен относительно x. А теперь рассмотрим «супер-универсум» возможных миров. Пусть даны миры x и y, и y либо доступен для x, либо не доступен. Как только определено, какой мир доступен для другого, мы имеем то, что технически называется *структурой.* Пусть дано некоторое предложение p и некоторый мир x, p либо истинно, либо ложно в x. Как только определено, какие предложения истинны и в каких мирах структуры они таковы, мы имеем то, что называется *моделью Крипке.* Следует понимать, что $\perp$ ложно в каждом из миров, и что $p \supset q$ истинно в мире, если и только если не тот случай, что p истинно, а q ложно в x. Таким образом, для каждого мира x множество всех предложений, истинных в x, принадлежит к типу 1. Для завершения описания следует сказать, что предложение Np объявляется истинным *в мире* x, если и только если, p истинно во всех мирах, которые *достижимы из x*. Мы будем говорить, что p *установлено* в модели, или *справедливо* в модели, если p истинно во *всех* мирах модели.

Полученное нами сейчас совпадает с тем, что было предметом обсуждения в последней главе, за исключением терминологии. Взамен элементов универсума, называемых *Мыслителями*, мы сейчас имеем *миры*. Вместо отношения «y есть родитель x» или «y есть предок x» мы сейчас говорим «y достижим от x». Наконец, вместо «Мыслитель x верит в p» мы сейчас имеем «p истинно в мире x». Имея в виду эти трансформации, все результаты прошлой главы получаются довольно легко. Множество всех предложений, которые справедливы в модели Крипке, должно быть типа 3 (Проблема 1, глава 22), и отсюда, модальная система K применима ко всем моделям Крипке.

Предположим, что мы сейчас добавляем условие транзитивности – для некоторых миров x, y и z, если y достижим от x и z достижим от y, тогда z достижим от x. Мы тогда имеем то, что называется *транзитивной* моделью Крипке. Ну, для любой транзитивной модели Крипке, множество всех предложений, которые истинны для всех миров модели, должно быть типа 4 (Проблема 2, глава 22), и поэтому подходящая модальная система в этом случае применима.

Таким образом, модальная система K применима для всех моделей Крипке, и модальная система K_4 применима ко всем *транзитивным* моделям Крипке. Эти два результата известны как теоремы семантической обоснованности для K и K_4. Крипке также доказал обратный результат: (1) если p справедливо во всех моделях Крипке, тогда p доказуемо в модальной системе K; (2) если p справедливо во всех *транзитивных* моделях Крипке, тогда p доказуемо в K_4. (Мы позднее объясним более точно, что имеется в виду под *доказуемостью* в модальной системе.) Эти два результата известны под названием *теорем полноты* для K и K_4.

Давайте будем говорить, что модель Крипке является *терминальной*, если справедливы следующие условия: при заданном некотором мире x модели, если мы переходим к миру x', достижимому от x, и затем к миру x'', достижимому от x', и, продолжая тот же процесс, приходим к миру y, от которого не достижим ни один мир (так называемые *терминальные* миры, которые подобны Мыслителям *без предков* из последней главы). Мы будем говорить, что модель принадлежит типу G, если она транзитивна и терминальна. В полным согласии с изложенным в последней главе мы видим, что класс предложений, который справедлив в модели типа G, должен быть типа G, и отсюда, подходящая модальная система это модальная система G. Мы, таким образом, получили так называемую теорему адекватности (soundness) для модальной системы G: каждое предложение, доказуемое в G, справедливо во всех транзитивных терминальных моделях. Обратное утверждение – теорема полноты для G – было установлено логиком Кристером Сегербергом. Оно говорит, что все предложения, которые справедливы во всех моделях типа G, доказуемы в модальной системе G[1].

В качестве философского анализа понятия необходимости модальная система G кажется в высшей степени неподходящей. Ее реальная важность лежит в области интерпретации доказуемости, к обсуждению которой мы переходим в следующей главе.

Система S_4 Льюиса. Льюис предложил несколько других систем модальной логики, одну из которых мы кратко упомянем. Льюис считал, что любое предложение, которое необходимо истинно, должно быть также истинным. (В лейбницевских терминах, если предложение истинно во

---

[1] Доказательства теорем полноты для модальных логик K, K_4, и G могут быт найдены в работе: *Boolos G.* The Logic of Provability (Cambridge University Press, 1993), и весьма упрощенное доказательство для G может быть найдено в книге: *Boolos G. Jeffrey C.* Computability and Logic. Cambridge University Press, 1980.

всех возможных мирах, оно должно быть определенно истинно в действительном мире!) И поэтому Льюис добавил в качестве аксиомы к K_4 все предложения формы $NX \supset X$. Эта система известна под названием *модальной системы S_4*.

Подходящей модельной теорией для S_4 является транзитивная модель Крипке (но не терминальная!) с дополнительным условием, что каждый мир достижим от самого себя. Легко видеть, что $NX \supset X$ справедливо в такой модели.

Модальные системы S_4 и G знаменуют истинное расхождение в модальной логике. Невозможно скомбинировать две системы в единую систему без получения противоречия (может ли читатель увидеть, почему?). И поэтому мы должны делать выбор, который зависит от наших целей. В качестве философского анализа понятия необходимой истины кажется вполне походящей система S_4. Для теории доказательства важной является система G. Но об этом будет следующая глава.

Упражнение 1. Почему невозможно скомбинировать системы G и S_4 без получения противоречия?

Упражнение 2. В модели типа G *никакой* мир не достижим от самого себя. Почему это так?

Упражнение 3. Доказать, что в модели типа G имеется по крайней мере одно предложение p и по крайней мере один мир x такие, что предложение $Np \supset p$ *ложно* в мире x.

Упражнение 4. Является ли следующее утверждение истинным или ложным? В модели Крипке типа G имеется по крайней мере один мир, в котором сумасшедшее утверждение $N\bot$ ($\bot$ необходимо истинно) на самом деле истинно.

ГЛАВА 24. *От необходимости к доказуемости*

Следующее важное развитие модальной логики после модальной семантики Крипке имело место в 1970-х, когда слову «необходимо» была придана *интерпретация доказуемости*. Удивительно, что этого не случилось раньше, так как Гедель предложил это в очень короткой статье, опубликованной в 1933 г. Гедель использовал символ «B» вместо льюисовского «N» и предположил, что Bp следует интерпретировать как p *доказуемо* (в системе Арифметики или же в любой из близко соотносящихся систем, исследованных Геделем). Все эти системы принадлежат типу 4, и поэтому аксиомы K_4 правильны при такой интерпретации. Однако математические системы оказались даже системами типа G (как это обнаружил Леб), отсюда, вполне естественным было изобрести аксиоматическую систему модальной логики для исследования таких систем. Так родилась система G. Она изучалась рядом логиков: включая Клаудио Бернарди, Джорджем Булосом, Д.Г.Дж. Ингом, Роберто Магари, Франко Монтана, Джиованни Самбин, Кристером Сегербергом, К. Сморински и Робертом Солове́й. Исследования этой системы все еще продолжаются.

На данном этапе полезно обсудить модальные аксиоматические системы более строго. Символизм модальной логики – это символизм пропозициональной логики, с одним добавленным новым символом – мы используем символ «B». (Вспомним, что Льюис использовал символ «N», а более распространенный символ – это «□». Но я предпочитаю геделевский символ «B».)

Под *модальной формулой* – более кратко, формулой – мы имеем в виду любое выражение, образованное по следующим правилам:

(1) $\perp$ есть формула, и каждая пропозициональная переменная p, q, r... есть формула.

(2) Если X и Y есть формулы, формулой является и $(X \supset Y)$.

(3) Если X есть формула, формулой является и BX.

То, что мы называли в главе 6 *формулой*, может сейчас быть названо *пропозициональной формулой*. Пропозициональная формула есть частный случай модальной формулы; это тот случай, когда в формулу не входит символ «B». Но мы будем иметь дело только с модальными формулами, и будем называть их просто *формулами*.

Логические связки $\neg$, $\vee$, $\supset$, $\equiv$ определяются с помощью $\supset$ и $\perp$ способом, который объяснялся в главе 8.

Каждая модальная система имеет свои собственные аксиомы. В каждой модальной системе, которые мы будем рассматривать, следует начинать с аксиом и последовательно переходить к доказательству новых формул с помощью следующих двух правил:

Правило 1 (известное как *modus ponens*). Имея доказанные X и $X \supset Y$, мы можем вывести Y.

Правило 2 (известное как вынуждение). Имея доказанное X, мы можем вывести BX.

Под *формальным доказательством* в системе подразумевается конечная последовательность формул (обычно представляемой вертикально и читаемая сверху вниз), называемая строчками доказательства, такая, что каждая строчка доказательства есть аксиома системы, или же получается из двух предшествующих строчек по Правилу 1, или же получается из предыдущей строчки по Правилу 2. Формула X называется *доказуемой* в системе, если имеется формальное доказательство, последней строчкой является X.

Три системы, которые представляют для нас конкретный интерес, это системы K, K_4 и G, чьи аксиомы мы рассмотрим ниже.

Аксиомы системы K:

(1) Все тавтологии.
(2) Все формулы формы $B(X \supset Y) \supset (BX \supset BY)$.

Аксиомы системы K4:
Аксиомы системы K вместе с

(3) Все формулы формы $BX \supset BBX$.

Аксиомы системы G:
Аксиомы системы K_4 вместе с

(4) Все формулы формы $B(BX \supset X) \supset BX$.

Замечание: Давайте называть аксиомы группы (4) специальными аксиомами для G. Мы вспомним теорему Крипке – де Йонга – Самбина из Главы 18, суть которой состоит в том, что если система типа 3 может доказать все предложения формы $B(BX \supset X) \supset BX$, тогда она может доказать также все предложения формы $BX \supset BBX$. Таким образом, в качестве альтернативы мы могли бы взять для G наши аксиомы из (1), (2) и (4); формулы группы (3) могли бы быть тогда выводными. Другими словами, если мы добавим аксиомы (4) к аксиомам системы K, а не к аксиомам K_4, мы все еще имели бы полную модальную систему G. Система G часто представляется именно в такой альтернативной форме.

Обсуждение. Знание этих модальных систем обеспечивает информацию о более привычных системах математики. Модальная система K является подходящей для системы S типа 3 (если мы интерпретируем BX как «X доказуемо в S»). Подобным же образом, аксиоматическая система K_4 хороша для системы S типа 4, а аксиоматическая система G справедлива для любой системы S типа G. Таким образом, эти модальные системы дают полезную информацию о доказуемости в более распространенных типах математических систем (которые не являются модальными системами). Компьютерные исследования в настоящее время также нуждаются в системах модальной логики по следующим причинам. Вообразим компьютерную программу, которая печатает различные предложения, некоторые из которых есть утверждения о том, что компьютер может печатать и чего он не может печатать. Тогда интерпретация BX становится следующей: «Компьютер может печатать X». Такие компьютеры являются, так сказать, «само-референтными», и следовательно, представляют интерес для тех, кто занимается проблемами искусственного интеллекта. Мы рассмотрим такие системы позднее.

Во многих случаях в этой книге я трактовал «верить» как модальность. Мы начинали с использования «B» как «верит» (Мыслитель подходящего типа верит...). Модальная логика позволяет нам дать унифицированную трактовку Мыслителей, которые *верят* в предложения, компьютеров, которые *печатают* предложения, и математических систем, которые *доказывают* предложения.

Сентенциональные модальные системы. Под модальным *предложением* мы имеем в виду модальную формулу, в которую не входит ни одна из пропозициональных переменных p, q, r – выражения типа $B\bot\supset\bot$, $B(\bot\supset B\bot)$. Таким образом, модальные предложения строятся из пяти символов B, $\bot$, $\supset$, (,). Под *сентенциональной* модальной системой мы подразумеваем модальную систему, чьими аксиомами являются все предложения (из чего легко следует, что доказуемы только предложения). Для любой модальной системы M мы обозначим через M* те системы, чьи аксиомы состоят из всех *предложений*, которые являются аксиомами M, и чьи правила вывода те же самые, что и у M (обычно это modus ponens и вынуждение). Мы будем, в частности, заинтересованы в сентенциональных модальных системах K*, K_4* и G*. Нетрудно показать, что если M является одной из трех этих систем, тогда любое *предложение*, доказуемое в системе M, также доказуемо в M*. (Читатель может попытаться сделать это в качестве упражнения – решение будет приведено позднее в главе 27, – когда будет нужна для обсуждения этого факта).

Мы вернемся к изучению модальных систем после рассмотрения восхитительной темы – самореференции – в следующей главе.

ЧАСТЬ X

СУТЬ ДЕЛА

ГЛАВА 25. *Геделизированная вселенная*

Давайте обратимся к тому, что может быть названо сутью дела, а именно – самореференцией. Мы все еще не дали читателю никакого представления о том, *как* Гедель ухитрился сконструировать самореферентное предложение – предложение, которое утверждает собственную недоказуемость в рассматриваемой системе. Он сделал это, изобретя удивительно остроумную процедуру, известную как *диагонализация.* В этой и следующих главах мы рассмотрим диагональный аргумент Геделя в нескольких формах.

Геделевский универсум

Давайте рассмотрим универсум, в котором есть бесконечное число Мыслителей. Есть также бесконечное число предложений об этом универсуме. Более точно:

(1) $\perp$ есть одно из этих предложение (и оно, конечно, ложно).

(2) Для любого из этих предложений p и любого Мыслителя R предложение, что R верит в p, есть одно из этих предложений.

(3) Для любых предложений p и q об универсуме $p \supset q$ есть опять-таки предложение об универсуме, которое истинно, если и только если либо p ложно, либо q истинно.

Следовательно, мы используем слово «предложение» как означающее предложение об универсуме. Мы определяем логические связки $\neg$, &, $\vee$, $\equiv$ через связки $\supset$ и $\perp$ тем же самым образом, как это делалось в главе 8.

Для любого Мыслителя R и любого предложения p, пусть Rp будет предложением, что R верит в p. Для любых Мыслителей R и S и любого предложения p, RSp есть предложение, что R верит, что S верит в p. Если имеется и третий Мыслитель K, KRSp есть предложение, что K верит, что R верит, что S верит в p. И подобным образом мы можем добавлять других Мыслителей.

Мыслители размышляли над этими предложениями, но все это было хаотичным, пока не появился некий логик из другого универсума и не навел порядок. Первое, что он заметил, что Мыслители не имеют имен, и он приписал каждому Мыслителю число (целое положительное число), известное под названием *геделева* числа Мыслителя. Никакие два Мыслителя не имеют одного и того же геделева числа, и каждое число есть число некоторого Мыслителя. Итак, Мыслители имеют имена: R_1 есть Мыслитель с числом 1, R_2 есть Мыслитель с числом 2, и так далее, и для каждого n R_n есть Мыслитель, чье число есть n.

Далее, логик построил все предложения об универсуме в определенной бесконечной последовательности $p_1, p_2, \ldots, p_n, \ldots$ Для каждого n число n известно как геделево число предложения p_n. После совершения всех этих действий логик вернулся в свой родной универсум.

Вскоре после отбытия логика наиболее проницательные Обитатели осознали следующий любопытный факт

Факт 1. Для каждого Мыслителя R имеется Мыслитель R* такой, что для любого предложения p_i, Мыслитель R* верит в p_i, если и только если, Мыслитель R верит, что R_i верит в p_i. (Таким образом, для любого Мыслителя R имеется Мыслитель R* такой, что для каждого числа i предложение $R^*p_i \equiv RR_ip_i$ истинно.)

Этот факт имеет некоторые интересные следствия, как это видно из следующих ниже проблем.

1. Диагональный принцип

Доказать, что для любого Мыслителя R имеется по крайней мере одно предложение p такое, что $p \equiv Rp$ истинно. (Другими словами, p истинно, если и только если, R верит в p.)

Доказательство: Рассмотрим Мыслителя R. В соответствии с Фактом 1 имеется Мыслитель R* такой, что для каждого числа i Мыслитель R* верит в p_i, если и только если, R верит в R_ip_i. Пусть R* имеет геделев номер h, и поэтому R* есть Мыслитель R_h. Таким образом, для любого числа i справедливо следующее:

(1) R_h верит в p, если и только если, R верит, что R_i верит в p_i.

Так как это истинно для *любого* числа i, это истинно и для числа h. Таким образом, мы имеем:

(2) R_h верит в p_h, если и только если, R верит, что R_h верит в p_h.

Если подставить букву p вместо предложения, что R_h верит в p_h, тогда получим $p \equiv Rp$.

2. Глупые Мыслители

Мыслитель *полностью глуп*, если он верит во все ложные предложения и не верит в истинные.

Доказать, что в этом универсуме не существует ни одного полностью глупого Мыслителя.

Доказательство: Мы только что видели, что у любого Мыслителя R имеется предложение p такое, что p истинно, если и только если, R верит в p. Это означает, что должен быть справедлив один из следующих случаев: (1) p истинно, и R верит в p; (2) p ложно, и R не верит в p. Если справедливо (1), тогда R верит в по крайней мере одно истинное предложение, а именно в p, и отсюда, R не полностью глуп. Если справедливо (2), тогда имеется по крайней мере одно ложное предложение, а именно p такое, что R не верит в p, и отсюда, R не верит ни в одно ложное предложение, и поэтому опять-таки R не полностью глуп.

После отбытия логика был обнаружен другой любопытный факт об этом универсуме:

Факт 2. Для любого Мыслителя R и любого предложения q имеется Мыслитель S такой, что для любого предложения p, S верит в p, если и только если, R верит в $p \supset q$. (То есть $p \equiv R(p \supset q)$ истинно.)

3. Принцип Тарского

Мыслитель этого универсума называется *совершенным*, если он верит во все истинные предложения и не верит ни в одно ложное. (Оно полностью противоположно полностью глупому Мыслителю.)

Целые годы ушли у Мыслителей этого универсума в поисках совершенного Мыслителя, но так и не смогли обнаружить такового. Почему их поиски были напрасными?

Решение: Используя Факт 2, мы берем в качестве q предложение $\perp$. Тогда для любого Мыслителя R, имеется Мыслитель R' (названный «S» в Факте 2), который верит в те и только те предложения p такие, что R верит в $p \supset \perp$. (Такой Мыслитель R' может полагаться *противоположным* R.)

Теперь предположим, что R совершенен. Тогда для любого предложения p, R верит в $p \supset \perp$, если и только если, $p \supset \perp$ истинно, которое, в свою очередь, истинно, если и только если, p ложно. Следовательно, R верит в $p \supset \perp$, если и только если, p ложно. Также R' верит в p, если и только если, R верит в $p \supset \perp$, Складывая эти два факта вместе, получаем, что R' верит в p, если и только если, p ложно. Это означает, что R' полностью глуп.

Мы, таким образом, видим, что если универсум содержит совершенно глупого Мыслителя R, тогда он должен содержать также совершенно глупого Мыслителя R'. Но мы доказали в Проблеме 2, что универсум не содержит совершенно глупого Мыслителя. Следовательно, универсум не содержит совершенно глупого Мыслителя.

Следующая важная вещь относительно этого универсума состоит в том, что определенные предложения называются *установленными,* и имеется Мыслитель E, который верит только в те и только те предложения, которые установлены.

4.

Предполагая, что все установленные предложения истинны, доказать, что множество установленных предложений неполно, то есть должно быть, по крайней мере, такое p, что не может быть установлено ни p, ни его отрицание. (Это означает также, что Мыслитель E никогда не сможет поверить в p и не сможет поверить в ¬p. Он вовеки останется неспособным разрешить вопрос, является p истинным или нет.)

Решение: Давайте назовем Мыслителя точным, если он не верит в ложные предложения. Рассмотрим сейчас некоторого точного Мыслителя R. Мы видели в Проблеме 3, что никакой Мыслитель не является совершенным, отсюда, R также несовершенен. Это означает, что R либо верит в некоторое ложное предложение, или же он не может поверить в некоторое истинное предложение. Так как R точен, он не верит ни в какое ложное предложение, отсюда должно быть, что R не может поверить в некоторое истинное предложение. Это доказывает, что для любого точного Мыслителя имеется, по крайней мере, одно истинное предложение, в которое он не может верить. Так как p истинно, ¬p ложно, отсюда R, будучи точным, не верит и в ¬p. Следовательно, для каждого точного Мыслителя R его система вер неполна. Имеется, по крайней мере, одно предложение p такое, что R не верит ни в p, ни в ¬p – для него остается навеки неразрешенным вопрос, является p истинным или ложным.

Предполагая, что все установленные предложения истинны, Мыслитель E точен (потому что он верит во все установленные предложения и ни в какие другие). Следовательно, имеется предложение p такое, что E не верит ни в p, ни в ¬p, отсюда, ни p и ни ¬p не являются установленными.

Далее, были обнаружены следующие факты.

Факт I. Для любого Мыслителя R имеется Мыслитель R* такой, что для каждого числа i предложение $R^*p_i \equiv RR_ip_i$ является *установленным*. (Это отличается от Факта 1 тем, что здесь мы говорим об *установленности* взамен *истинности*.)

Факт II. Для любого Мыслителя R и любого предложения q имеется Мыслитель S такой, что для каждого числа i $Sp_i \equiv R(p_i \supset q)$ установлено. (Это отличается от Факта 2 в том отношении, что здесь мы говорим об установленности взамен истинности.)

5.

Доказать, что для любого Мыслителя R имеется предложение p такое, что предложение $p \equiv Rp$ установлено.

Доказательство: Доказательство в существенных чертах совпадает с доказательством Проблемы 1, используя Факт I вместо Факта 1.

Пусть дан Мыслитель R, имеется Мыслитель R_h (называемый R*) такой, что для любого числа I предложение $R_hp_i \equiv RR_ip_i$ является установленным. Отсюда, установленным является и $R_hp_h \equiv RR_hp_h$. Таким образом, $p \equiv Rp$ является установленным, где p есть предложение R_hp_h.

6.

Предположим, что множество установленных предложений принадлежит к типу 1.

(a) Показать, что для каждого Мыслителя R и каждого предложения q, имеется предложение p такое, что предложение $p \equiv R(p \supset q)$ является установленным.
(b) Показать, что для каждого Мыслителя R и каждого предложения q, имеется предложение p такое, что предложение $p \equiv (Rp \supset q)$ является установленным.

Решение: Предположим, что множество установленных предложений принадлежит типу 1.

(a) Возьмем некоторого Мыслителя R и некоторое предложение q. По Факту 2, имеется Мыслитель S такой, что для каждого p предложение $Sp \equiv R(p \supset q)$ является установленным. Также, по Проблеме 5 (беря вместо буквы R букву S), имеется предложение p такое, что $p \equiv Sp$ является установленным. Отсюда следует, что установленным является и предложение $p \equiv R(p \supset q)$ (так как оно является логическим следствием двух последних предложений).

(b) Возьмем опять некоторого Мыслителя R и некоторое предложение q. Согласно (a), имеется предложение – назовем его p_1 – такое, что $p_1 \equiv (p_1 \supset q)$ является установленным. Отсюда, $(p_1 \supset) \equiv (R(p_1 \supset q) \supset q)$ является установленным (трюк, который мы уже использовали), и поэтому установленным является и предложение $p \equiv (Rp \supset q)$, где p есть предложение $p_1 \supset q$.

Наконец, был осознан и следующий факт.

Факт III. Мыслитель E принадлежит к типу 4 (и отсюда, множество установленных предложений принадлежит типу 4).

7.

Доказать, что множество установленных предложений принадлежит типу G.

Доказательство: Нам сейчас дано, что E принадлежит типу 4, отсюда, он определенно принадлежит типу 1. Тогда, согласно (b) из последней Проблемы, для любого предложения q имеется предложение p такое, что предложение $p \equiv (Ep \supset q)$ является установленным. Таким образом, множество установленных предложений является *рефлексивным*. А мы знаем из главы 19, что любая рефлексивная система типа 4 принадлежит также к типу G.

Мы сейчас видим, что множество установленных предложений этого универсума принадлежит к типу G, и отсюда, если это множество непротиворечиво, тот самый *факт*, что оно непротиворечиво и истинно, не является одним из установленных предложений в этом универсуме. Эквивалентно, если Мыслитель E непротиворечив, он никогда не узнает, что он непротиворечив.

Отношение к математическим системам

Читатель может удивиться, а какое отношение все это имеет к теории математических систем. Ну, предположим, мы имеем математическую систему S со всеми ее предложениями, построенными в некоторую бесконечную последовательность $p_1, p_2, \ldots, p_n, \ldots$ Теперь, взамен *Мыслителей* мы рассмотрим определенные *свойства* предложений; эти свойства также построены в некоторую бесконечную последовательность $R_1, R_2, \ldots, R_n, \ldots$ Для любого *свойства* R_i и любого предложения p_j, R_ip_j будет предложение, что свойство R_i справедливо для предложения p_i. Предположим также, что определенное свойство – назовем его E – быть *доказуемым* предложением системы – это одно из свойств из перечня выше, и предположим, что Факты 1, 2 и 3 справедливы при замене *Мыслителя* на *свойство* и *установлен* – на *доказуемо* (то есть, доказуемо в S). Тогда простым изменением терминологии предшествующие аргументы этой главы показывают, что система S должна быть системой типа G.

На самом деле, системы, исследованные Геделем, начинаются не со свойств предложений, а со свойств *чисел*. Но с помощью геделевской нумерации предложений, любое свойство чисел соответствует определенному свойству предложений, а именно, для некоторого свойства A *чисел*, пусть A' будет свойством, которое справедливо точно для *предложений* p_i таких, что A справедливо для числа i. Конкретная иллюстрация этого будет дана в следующей главе.

Самореференция может быть достигнута и без нумерации Геделя, как показано в упражнении ниже.

Упражнение 1. Имеется другой универсум Мыслителей, в которых нет различия между тем, конечное или бесконечное число этих Мыслителей. Некоторые из Мыслителей бессмертны, но ни один из Мыслителей не знает, какой из них бессмертен, а какой смертен. На самом деле, ни один Мыслитель не знает даже про себя, смертен он или бессмертен. Для каждого Мыслителя R пусть $\overline{R}$ будет предложением, что R бессмертен. Для любых Мыслителей R и S пусть $R\overline{S}$ будет предложением, что R *верит,* что S бессмертен; для любых трех Мыслителей R, S и K пусть $RS\overline{K}$ будет предложением, что R верит, что S верит, что K бессмертен, и т.д. (если в рассмотрение включено больше Мыслителей).

Вместо Факта 1 из последнего универсума мы имеем в виду следующий факт для этого универсума: для любого Мыслителя R имеется Мыслитель R* такой, что для каждого Мыслителя S Мыслитель R* верит, что S бессмертен, если и только если, R верит, что S сам верит в собственное бессмертие (таким образом, $R^*\overline{S}$ истинно, если и только если, $RS\overline{S}$ истинно).

Пусть задан Мыслитель R, найти предложение p такое, что p истинно, если и только если, R верит в p.

Замечание. Этот метод получения самореференции без геделевых чисел заимствован из области, известной под названием комбинаторной логики. Масса проблем с самореференцией может быть найдена в моей книге «Передразнить пересмешника» (To Mock a Mockingbird).

ГЛАВА 26. *Некоторые замечательные логические машины*

Самая последняя машина Фергюссона

Логик Малькольм Фергюсон из моей книги «*Принцесса или Тигр*» любил строить логические машины для иллюстрации важных принципов логики и теории доказательств. Одна из его машин описана в этой книге. Когда в последние годы Фергюссон услышал о теоремах Геделя и Лёба, он немедленно приступил к конструированию второй машины, которую с восторгом демонстрировал своим друзьям. К их удовлетворению, он доказал, что устройство было непротиворечивой и стабильной машиной типа G, и особый его восторг при демонстрации вызывало то, что машина, будучи непротиворечивой, тем не менее никогда не могла доказать свою собственную непротиворечивость! Машина иллюстрировала в очень простой и поучительной манере существенные идеи, которые лежат в основе Первой и Второй теорем Геделя о неполноте, а также теоремы Лёба. Соответственно я с удовольствием сообщаю детали этих вещей читателю.

Машина печатает предложения, состоящие из 17 символов. Первые семь символов таковы:

$$\begin{array}{ccccccc} P & \perp & \supset & (&) & d & , \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array}$$

Под каждым из этих символов я подписал их *геделевы номера*. Остальные десять символов – это 1, 2, 3, ..., 0. Этим цифрам приписываются геделевы номера следующим образом: геделев номер цифры 1 есть 89 (за 8 следует 9); геделев номер цифры 2 есть 899 (за 8 следует две 9); и т.д., вплоть до цифры 0, чей геделев номер есть 89999999999 (за цифрой 8 следует десять цифр 9). Таким образом, каждый из семнадцати символов имеет свой геделев номер. Если дано сложное выражение, его геделев номер находится подстановкой вместо каждого символа его геделева номера, например геделев номер выражения (P⊥⊃⊥) есть 412325. Другой пример: геделев номер выражения P35 есть 1899899999. Для любого выражения E, под $\overline{\text{E}}$ мы подразумеваем геделевский номер E (записанное в виде цепочки из цифр 1, 2, ..., 0). Не каждое число есть геделев номер выражения (например, 88 не является геделевым номером никакого выражения). Если n есть геделев номер некоторого выражения, мы иногда говорим об этом выражении как об n-м выражении. (Например, pd есть 16-е выражение; ⊥ есть 2-е выражение.)

Машина является *самореферентной* в том отношении, что предложения, которые печатаются машиной, выражают предложения о том, что машина может и чего не может печатать машина. Выражение называется *печатаемым*, если машина может напечатать его. Символ «P» означает «печатаемое», и для любого выражения E, построенного из семнадцати символов, если мы хотим выписать предложение, которое утверждает, что E печатаемо, мы пишем не PE, а P$\overline{\text{E}}$ (то есть P, за которым следует геделев номер E).

Например, предложение, которое утверждает, что $(P\bot\supset\bot)$ печатаемо, есть $P(\overline{P\bot\supset\bot})$, то есть P412325.

Для любых двух выражений X и Y Фергюссон определяет *диагонализацию X в отношении Y* как выражение $(X(\overline{X}, \overline{Y})\supset Y)$. Если принять для термина «диагонализация» сокращение d, тогда мы имеем $Pd(\overline{X}, \overline{Y})$ как предложение, выражающее суждение, что диагонализация X в отношении Y печатаема.

Мы сейчас определим, что означает для выражения быть *предложением* и что означает для предложения быть *истинным*.

1. $\bot$ есть предложение, и $\bot$ ложно.
2. Для любого выражения X выражение $P\overline{X}$ есть предложение, и оно истинно, если и только если, выражение X печатаемо.
3. Для любых выражений X и Y выражение $Pd(\overline{X}, \overline{Y})$ есть предложение, и оно истинно, если и только если, выражение $(X(\overline{X}, \overline{Y})\supset Y)$ – которое есть диагонализация X в отношении Y – печатаемо.
4. Для любых двух предложений X и Y, выражение $(X\supset Y)$ есть предложение, и оно истинно, если и только если, X не истинно или же Y истинно.

Следует понимать, что никакое выражение не есть предложение, если оно не есть следствие правил, приведенных выше. Логические связки $\neg$, &, $\vee$ и $\equiv$ определяются с помощью $\supset$ и $\perp$ тем способом, который описан в главе 8.

Теперь мы можем дать правила для того, что может печатать машина. Машина запрограммирована последовательно печатать бесконечный перечень предложений. Некоторые предложения, называемые *аксиомами*, могут быть напечатаны на любой стадии процесса. Все аксиомы есть тавтологии. (Таким образом, для любой тавтологии X машина может напечатать X в любое время, независимо от того, печатала или нет она это выражение ранее.) Далее, машина запрограммирована так, что для любых двух предложений X и Y, если машина на определенной стадии напечатала X и X$\supset$Y, тогда она может напечатать Y. Таким образом, это машина типа 1 (в смысле, что класс печатаемых предложений принадлежит типу 1). Так как истинно, что если X и X$\supset$Y оба печатаемы, печатаемо и Y, тогда предложение $(P\overline{X} \,\&\, P(\overline{X \supset Y}) \supset P\overline{Y}$ истинно, или, что то же самое, истинно предложение $P(\overline{X \supset Y}) \supset (P\overline{X} \supset P\overline{Y})$. (Эти два предложения логически эквивалентны). Ну, машина «знает» истинность всех предложений формы $P(\overline{X \supset Y}) \supset (P\overline{X} \supset P\overline{Y})$ и берет их в качестве аксиом. Таким образом, это машина типа 2. Далее, если машина когда-либо печатает предложение X, она «знает», что напечатала X, и рано или поздно напечатает истинное предложение $P\overline{X}$. (Предложение $P\overline{X}$ истинно, так как X напечатано). И поэтому машина нормальна, и отсюда, типа 3. Так как машина нормальна, для любого предложения X предложение $P\overline{X} \supset P(\overline{P\overline{X}})$ истинно. Поэтому машина исходно «осознает» истинность всех таких предложений и рассматривает их как аксиомы. Таким образом, это машина типа 4.

Есть еще одна вещь, которую может делать машина, и это является очень важным. Для любых двух выражений X и Y предложение $Pd(\bar{X}, \bar{Y})$ истинно, если и только если, $(X(\bar{X}, \bar{Y}) \supset Y)$ печатаемо, которое, в свою очередь имеет место, если и только если, предложение $P(\overline{\bar{X}(\bar{X}, \bar{Y}) \supset \bar{Y}})$ истинно. Значит, следующее предложение истинно: $Pd(\bar{X}, \bar{Y}) \equiv P(\overline{X(\bar{X}, \bar{Y}) \supset Y})$. Итак, машина знает истинность всех таких предложений и рассматривает их как аксиомы. Они называются *диагональными аксиомами*.

Перейдем к систематическому рассмотрению всех аксиом и операций машины.

Аксиомы:

Группа 1. Все тавтологии.

Группа 2. Все предложения формы $P(\overline{X \supset Y}) \supset \overline{(P\bar{X} \supset P\bar{Y})}$.

Группа 3. Все предложения формы $P\bar{X} \supset P(\overline{P\bar{X}})$.

Группа 4 (диагональные аксиомы). Все предложения формы $Pd(\bar{X}, \bar{Y}) \equiv P(\overline{X(\bar{X}, \bar{Y})} \supset Y)$, где X и Y являются любыми выражениями (не обязательно предложениями).

Операционные правила:

1. Аксиомы могут быть напечатаны на любой стадии.
2. Если напечатаны предложения X и $(X \supset Y)$, тогда машина может напечатать Y.

3. Если напечатано предложение X, машина может напечатать $P\overline{X}$.

Это все правила, управляющие печатанием машины. Следует понять, что единственный способ, которым машина может напечатать предложение X на данной стадии, состоит в том, чтобы следовать этим правилам. То есть, X печатаема на данной стадии *только* в том случае, если соблюдаются следующие три условия: (1) X есть аксиома; (2) имеется предложение Y такое, что Y и ($Y \supset X$) оба были напечатаны на более ранней стадии; (3) имеется предложение Y такое, что X есть предложение $P\overline{Y}$ и Y было напечатано на более ранней стадии.

Замечание. Для каждого предложения X, пусть BX будет предложением $P\overline{X}$. Символ «B» *не* есть часть машинного языка; мы используем его для *разговора о* машине. Символ «B» стоит для операции, которая приписывает каждому предложению X предложение $P\overline{X}$. Когда мы говорим, что машина типа 4, мы имеем в виду, что она типа 4 со ссылкой на Операцию B. В сущности, без диагональных аксиом аксиоматическая система этой машины есть модальная система K_4. Мы вскоре увидим, что добавление диагональной аксиомы дает нам всю силу модальной системы G.

Доказуемость

Мы определили для каждого предложения, что для предложения означает быть *истинным*, и поэтому каждое предложение выражает определенное суждение, которое может быть истинным или ложным. Мы говорим, что машина *доказывает* суждение, если она печатает некоторое предложение, которое выражает суждение. Например, предложение $\neg P2$ выражает суждение, что машина непротиворечива (так как 2 есть геделев номер $\perp$), и поэтому если машина напечатала $\neg P2$, она докажет свою собственную *противоречивость*.

Мы будем говорить, что машина *точна*, если все предложения, доказуемые машиной, истинны. Мы будем говорить, что машина *непротиворечива*, если она не может доказать $\perp$, и что машина *устойчива*, если для каждого предложения X, если $P\overline{X}$ печатаемо, тогда печатаемо и X.

Рефлексивность

Теперь мы переходим к доказательству, что машина является геделевой, то есть рефлексивной.

1. Геделево предложение G

Найти предложение G такое, что предложение $G \equiv \neg P\overline{G}$, то есть предложение $G \equiv (P\overline{G} \supset \perp)$, печатаемо.

2. Рефлексивность

Показать, что для любого предложения Y имеется предложение X такое, что предложение $X \equiv (P\overline{X} \supset Y)$ печатаемо.

Решение: Проблема 1 есть частный случай Проблемы 2, поэтому сперва мы решим Проблему 2. Пусть Y будет некоторое предложение. Для данного выражения Z, предложение $Pd(\overline{Z}, \overline{Y}) \equiv P(\overline{Z(\overline{Z}, \overline{Y}) \supset Y})$ является печатаемым (потому что это одна из диагональных аксиом). Мы берем в качестве Z выражение Pd, и следовательно, $Pd(\overline{Pd, Y}) \equiv P(\overline{Pd(\overline{Pd}, \overline{Y}) \supset Y})$ печатаемо. Машина принадлежит к типу 1, отсюда следует, что следующее предложение печатаемо: $(Pd(\overline{\overline{Pd}, \overline{Y}}) \equiv (P(\overline{Pd(\overline{Pd}, \overline{Y}) \supset Y}) \supset Y)$. Таким образом, предложение $X \equiv (P\overline{X} \supset Y)$ печатаемо, где X есть предложение $(Pd(\overline{Pd}, \overline{Y}) \supset Y)$.

Проблема 1 есть специальный случай Проблемы 2, что очевидно, если мы возьмем вместо Y константу $\perp$. Таким образом, геделево предложение G для этой машины есть $Pd(\overline{Pd}, \overline{\perp}) \supset \perp$, то есть предложение $(Pd(16, 2) \supset \perp)$.

Рассмотрим более тщательно геделево предложение G. Что означает предложение (Pd(16,2)? Оно утверждает, что диагонализация 16-го выражения в отношении 2-го выражения печатаема. Но 16-е выражение есть Pd, а 2-е выражение есть $\perp$, и поэтому Pd(16,2) говорит, что диагонализация Pd в отношении $\perp$ печатаема, но эта диагонализация есть предложение $(Pd(16,2) \supset \perp)$, то есть само предложение G! И поэтому Pd(16,2) говорит, что G *печатаемо*, отсюда $(Pd(16,2) \supset \perp)$ – которое есть предложение G – говорит, что G *не* печатаемо (или, что то же самое, что печатаемость G влечет ложность). Таким образом, G говорит, что G не печатаемо; G *истинно*, если и только если, G не печатаемо. Таким образом, G утверждает свою собственную непечатаемость. В этом и состоит, вкратце, изобретательный метод Геделя в достижении самореференции.

Предложение $G \equiv \neg P\overline{G}$ то есть, предложение $G \equiv (P\overline{G} \supset \perp)$, не только истинно, но и в действительности печатаемо (это одна из диагональных аксиом). Так как машина нормальна и типа 1, по Первой Теореме о Неполноте Геделя (Теорема I, Глава 20) следует, что если машина непротиворечива, тогда G непечатаемо, и если машина устойчива, тогда $\neg G$ также непечатаемо. И поэтому, если машина непротиворечива и стабильна одновременно, предложение G неразрешимо в системе предложений, которые может печатать машина.

Теперь, машина на самом деле типа 4, и так как она геделева, предложение $G \equiv \neg P\overline{G}$ печатаемо – из Второй Теоремы о Неполноте Геделя (часть 4 Итогов I*, глава 13) следует, что если машина непротиворечива, тогда она не может доказать свою собственную непротиворечивость, то есть не может напечатать предложение $\neg P2$. Также, если машина непротиворечива, тогда предложение $\neg P2$ истинно и поэтому представляет другой пример истинного предложения, которое машина не может напечатать.

Далее, машина рефлексивна (Проблема 2) и, будучи типа 4, должна быть лёбовской (по Теореме Лёба), и поэтому для любого предложения X, если $P\overline{X} \supset X$ печатаемо, печатаемым будет и X. По Теореме M_1 (глава 18), машина в этом случае будет типа G.

Правильность машины

Мы показали, что *если* машина Фергюссона непротиворечива, тогда она не может доказать собственную непротиворечивость; но откуда мы знаем, непротиворечива машина или нет? Мы сейчас докажем, что машина не только непротиворечива, но полностью точна, то есть, что каждое предложение, которое печатает машина, истинно.

Мы уже показали, что все аксиомы машины истинны, но давайте тщательно рассмотрим аргументацию. Аксиомы группы 1 являются тавтологиями, отсюда, они определенно истинны. Что касается аксиом группы 2, то сказать, что $P(\overline{X \supset Y}) \supset (P\overline{X} \supset P\overline{Y})$ истинно, значит сказать, что если $P(\overline{X \supset Y})$ и $P\overline{X}$ истинны, истинным является и $P\overline{Y}$. Другими словами, если (X⊃Y) и X печатаемы, печатаемым является и Y. Ну, это ясно благодаря Операции 2. Таким образом, аксиомы Группы 2 истинны. Что касается аксиом Группы 3, то сказать, что $P\overline{X} \supset P(\overline{P\overline{X}})$ истинно, значит сказать, что если $P\overline{X}$ истинно, то истинным является и $P(\overline{P\overline{X}})$. Другими словами, если X печатаемо, то печатаемо и $P\overline{X}$, что и имеет место благодаря Операции 3. Что касается диагональных аксиом, $Pd(\overline{X}, \overline{Y})$ истинно, если и только если, $(X(\overline{X}, \overline{Y}) \supset Y)$ печатаемо, или истинно, если и только если $P(\overline{X, (\overline{X}, \overline{Y}) \supset Y})$ истинно. Следовательно, $Pd(\overline{X}, \overline{Y}) \equiv$ $\equiv P(\overline{X, (\overline{X}, \overline{Y}) \supset Y})$ истинно.

Сейчас мы знаем, что все аксиомы машины истинны, но нам еще нужно показать, что все печатаемые предложения истинны.

Мы вспомним, что машины печатают предложения на различных *стадиях*. Мы сейчас хотим установить следующие лемму, теорему и следствие.

Лемма. Если X есть предложение, напечатанное на определенной стадии, и все предложения, напечатанные до этой стадии, истинны, тогда X также истинно.

Теорема 1. Каждое предложение, напечатанное машиной, истинно.

Следствие. Машина непротиворечива и одновременно устойчива.

3.

Как доказать эти утверждения?

Решение: Сначала мы докажем Лемму: Предположим, что все до сих пор напечатанные предложения истинны. Нам нужно показать, что X истинно.

Случай 1. X есть аксиома. Тогда X истинно (как мы уже доказали).

Случай 2. Имеется предложение Y такое, что Y и $(X \supset Y)$ оба были уже напечатаны. Тогда по предположению, Y и $(X \supset Y)$ оба истинны, отсюда и X должно быть истинным.

Случай 3. X имеет форму $P\overline{Y}$, где Y есть предложение, которое было уже напечатано. Так как Y напечатано, $P\overline{Y}$ истинно, то есть X истинно.

Это завершает доказательство Леммы.

Доказательство Теоремы 1. Машина запрограммирована печатать все печатаемые предложения в некоторой определенной бесконечной последовательности X_1, X_2, ..., X_n, ... Под X_n подразумевается предложение, напечатанное на стадии n. Теперь, первое предложение, напечатанное машиной (предложение X_1), должно быть аксиомой (так как машина не может пока напечатать других предложений), отсюда X_1 должно быт истинным. Если приведенный выше бесконечный список содержит какое-либо ложное предложение, тогда должно быть *наименьшее* число n такое, что X_n ложно, то есть должно быть *первое* ложное предложение, которое напечатает машина. Мы знаем, что n не равно 1 (так как X_1 истинно), отсюда n должно быть больше 1. Это значит, что машина напечатает ложное предложение на стадии n, но она печатала только истинные предложения на более ранних стадиях. Но это противоречит Лемме. Следовательно, машина никогда не сможет напечатать никаких ложных предложений.

Доказательство Следствия 1: Так как машина точна (по Теореме 1), никогда не может быть напечатано $\perp$, потому что $\perp$ ложно. Следовательно, машина непротиворечива.

Далее, предположим, что $P\overline{X}$ печатаемо. Тогда $P\overline{X}$ истинно (по Теореме 1), что означает, что X печатаемо. Следовательно, машина устойчива.

Мы сейчас видим, что машина Фергюссона непротиворечива, но никогда не сможет доказать своей собственной непротиворечивости. Таким образом, вы и я (как и Фергюссон) знаем, что машина непротиворечива, но бедная машина не имеет этого знания.

Вариант Крейга

Когда инспектор Крейг (хороший друг Фергюссона) услышал о машине Фергюссона, он подумал об интересном варианте, который не включает геделевой нумерации. Машина Крейга имеет следующие шесть символов:

$$P \perp \supset (\ ) R$$

Его определения предложения и истинного предложения задаются следующими правилами:

(1) $\perp$ есть предложение, и $\perp$ не истинно.
(2) Для любых двух предложений X и Y, (X$\supset$Y) есть предложение, и оно истинно, если и только если, X не истинно или Y истинно.
(3) Для некоторого предложения X, PX истинно, если и только если, машина может напечатать X.
(4) Для любых предложений X и Y выражение (XRY) есть предложение, и оно истинно, если и только если, машина может напечатать ((XRY)$\supset$Y).

С первого взгляда может показаться, что (4) содержит порочный круг, так как истинность (XRY) определяется в терминах выражения, включающего букву R; но круг тут только кажущийся. Если бы Крейг определил (XRY) как истинное, если и только если, *истинно* ((XRY)$\supset$Y), определение было бы круговым, так как мы не могли бы знать, что значило бы для (XRY) быть истинным без предварительного знания, что значит быть истинным для (XRX). Но Крейг не делал этого. Факт в том, что для любых предложений X и Y выражение ((XRY)$\supset$Y) либо печатаемо, либо нет. Символ «R» стоит для отношения, которое содержится для между X и Y, когда ((XRY)$\supset$Y) *печатаемо*. Таким образом, Крейг не определял отношение R в терминах самого этого отношения, а определял его в терминах *символа* «R», а это уже не включает порочного круга.

Первые три группы аксиом машины Крейга те же самые, что и аксиомы для системы K_4 (где *предложение* означает предложение в машинном языке системы Крейга). Таким образом, они подобны первым трем группам аксиом машины Фергюссона, только без знака «-» над буквами. Четвертая группа аксиомы Крейга может читаться так:

(4') (Диагональная аксиома Крейга). Все предложения диагональной аксиомы истинны.

Операционные правила машины Крейга те же самые, что и для машины Фергюссона.

4.

(a) Доказать, что машина Крейга рефлексивна (и отсюда, принадлежит также типу G, так как она типа 4).

(b) Найти геделево предложение для машины Крейга (то есть предложение G такое, что $G \equiv \neg PG$ печатаемо машиной Крейга).

Решение: (a) Так как для любых двух предложений X и Y, предложение

$(XRY) \equiv P((XRY) \supset Y)$ печатаемо (это диагональная аксиомы), это означает, что X есть то же самое предложение Y. Следовательно, $(YRY) \equiv P((YRY) \supset Y)$ печатаемо, отсюда печатаемо и предложение $((YRY \supset Y) \equiv (P(YRY) \supset Y) \supset Y)$, и таким образом, печатаемо и $Z \equiv (PZ \supset Y)$, где Z есть предложение $((YRY) \supset Y)$.

(b) Взяв $\perp$ для Y, мы получаем геделево предложение $((\perp R \perp) \supset \perp)$.

Замечание. Все аксиомы машины Крейга также истинны, и из тех же самых соображений, которые использовались при обсуждении машины Фергюссона, видно, что каждое предложение, печатаемое машиной Крейга, истинно. Следовательно, машина Крейга также непротиворечива и устойчива (и принадлежит к типу G).

Наблюдение МакКаллоха

Когда Уолтер МакКаллох (друг Крейга и Фергюссона) узнал о машине Крейга, он сделал следующее интересное наблюдение: Пусть даны предложения X и Y, в которые не входит символ «R». Имеется предложение Z, в которое не входит «R», и такое, что предложение $(XRY \equiv Z)$ печатаемо. (Это влечет, между прочим, что для любого предложения X имеется предложение X', в которое не входит «R», такое, что $X \equiv X'$ печатаемо машиной Крейга.)

Можете ли доказать, что наблюдение МакКаллоха верно?

Решение: Мы доказали в решении Проблемы 8 из Главы 19, что любая система типа G, которая может доказать $p \equiv B(p \supset q)$, может доказать и $p \equiv Bp$. (Мы показали это для Мыслителей, но тот же самый аргумент подходит и для систем.) Теперь, система Крейга принадлежит типу G, и для любого предложения X предложение $(XRX) \equiv$ $\equiv P((XRX) \supset X)$ печатаемо, и поэтому, если мы возьмем (XRX) для p и X для q, получим, что $(XRX) \equiv PX$ печатаемо. Из этого следует, что $((XRX) \supset Y) \equiv (PX \supset Y)$ печатаемо, и отсюда (по свойству регулярности), $P((XRX) \supset Y) \equiv P(PX \supset Y)$ печатаемо. Но печатаемо и $(XRY) \equiv P((XRX) \supset Y)$, отсюда, печатаемо $(XRY) \equiv P(PX \supset Y)$. Если символ «R» не входит ни в X, ни в Y, тогда он не входит и в $P(PX \supset Y)$, и поэтому мы берем в качестве Z предложение $P(PX \supset Y)$.

ГЛАВА 27. *Самоприменимые модальные системы*

Вдохновленный логикой машин Крейга и Фергюссона, я хочу сейчас рассмотреть модальные аксиоматические системы с точки зрения *самореферентных интерпретаций.*

Вспомним, что под *модальным предложением* – более кратко, *предложением* – мы имеем в виду модальную формулу без пропозициональных переменных. Мы сейчас определяем модальное предложение как *истинное* для модальной системы M, если оно истинно, когда мы интерпретируем B как *доказуемое в M.* Таким образом:

(1) $\perp$ ложно для M.

(2) Для любых модальных предложений X и Y предложение $X \supset Y$ истинно в M, если и только если, либо X не истинно в M, или же Y не истинно в M.

(3) Для любого предложения X предложение BX истинно в M, если и только если X *доказуемо* в M.

Замечание: Предложение X может быть *истинным* в модальной системе M без того, чтобы быть доказуемым в M, и обратно. Например, предложение $\neg B\perp$ *истинно* в M, если и только если, M непротиворечиво. Сказать, что $\neg B\perp$ доказуемо в M, значит сказать, что M может доказать свою собственную непротиворечивость. Мы вскоре увидим, что модальная система G непротиворечива, и поэтому предложение $\neg B\perp$ *истинно* в G. Но предложение $\neg B\perp$ *не доказуемо* в G. С другой стороны, любая противоречивая модальная система типа 1 может доказать все предложения, и отсюда, предложение $\neg B\bot$. В этом случае предложение $\neg B\bot$ *доказуемо* в системе, но не *истинно* в системе.

Рассмотрим сейчас модальную систему M_1 и модальную систему M_2 (возможно, тождественную с M_1, а возможно и отличную от нее). Мы будем говорить, что M_1 правильно в M_2, если каждое предложение, доказуемое в M_1, истинно в M_2. И мы будем говорить, что модальная система M является самореферентно правильной, если M правильно в M – другими словами, если каждое предложение, доказуемое в M, истинно в M.

Любая самореферентно правильная система должна быть непротиворечивой (потому что если бы $\bot$ было доказуемо в системе, система не могла бы быть самореферентно правильной, так как $\bot$ ложно в этой системе), а также должна быть устойчивой (потому что если BX доказуемо в системе, а система референтно правильна, тогда BX должно быть истинно в системе, что означает, что X доказуемо в системе). И поэтому любая самореферентно правильная система автоматически становится одновременно непротиворечивой и устойчивой.

Самореферентная правильность имеет одну любопытную особенность. Возможно, что система M могла бы быть самореферентно правильной, и все же, если убрать некоторые аксиомы, получающаяся система больше не будет самореферентно правильной. Например, мы могли бы иметь одну аксиому, которая утверждает, что вторая аксиома доказуема в системе; если эту вторую аксиому убрать, первая аксиома становится ложной!

Некоторые самореферентно правильные системы. Мы вспомним сентенциональные модальные системы $\bar{K}$, $\bar{K}_4$ и $\bar{G}$, описанные в главе 24 (они подобны системам K,

K_4 и G, за исключением того, что все аксиомы ограничены предложениями). Мы сейчас ставим целью показать, что эти три системы являются самореферентно правильными (из чего, кстати говоря, мы сможем показать, что системы K, K4 и G являются самореферентно правильными).

Если $\overline{M}$ есть любая из трех аксиоматических систем $\overline{K}$, $\overline{K}_4$ и $\overline{G}$, чтобы показать, что $\overline{M}$ самореферентно правильна, достаточно показать, что все *аксиомы* $\overline{M}$ истинны в $\overline{M}$ – причина этого заключается в следствии следующей Леммы, которая имеет и другие применения.

Лемма А. Пусть $\overline{M}_1$ будет любой сентенциональной модальной системой, единственными правилами вывода являются *modus ponens* (из X и $X \supset Y$ мы можем вывести Y) и правило вынуждения (из X мы можем BX). Пусть M_2 будет любая модальная система такая, что все предложения, доказуемые в $\overline{M}_1$, доказуемы в M_2, и такая, что все *аксиомы* $\overline{M}_1$ *истинны* в M_2. Тогда все предложения, доказуемые в $\overline{M}_1$, истинны в M_2, то есть $\overline{M}_1$ правильна в M_2.

Следствие А1. Для любой сентенциональной модальной системы $\overline{M}$, чьими единственными правилами вывода являются *modus ponens* и вынуждение, если все *аксиомы* $\overline{M}$ истинны в $\overline{M}$, тогда $\overline{M}$ самореферентно правильна.

1.

Доказать Лемму А. (Подсказка: Доказательство в существенных чертах то же самое, что и в аргументе, который мы привели в последней главе для демонстрации того, что все доказуемые предложения машины Фергюссона истинны, – раз мы установили, что все аксиомы машины были истинны.)

Решение: Рассмотрим некоторую последовательность X_1, X_2, ..., X_n предложений, которые составляют доказательство в системе $\overline{M}_1$. Мы увидим, что первая строчка X_1 должна быть истинна в M_2, тогда вторая строчка X_2 должна быть истинна в M_2, тогда третья строчка X_3..., и так далее, вплоть до последней строчки.

Первая строчка X_1 должна быть аксиомой, отсюда она истинна в M_2 по предположению. Теперь рассмотрим вторую строчку X_2. Либо это аксиома $\overline{M}_1$ (в этом случае она истинна в M_2), или же она должна предложением BX_1, и в этом случае она определенно истинна в M_2, так как X_1 уже было доказано в M_1, и отсюда, доказуемо в M_2. Сейчас мы знаем, что первые две строчки истинны в M_2. Мы сейчас рассматриваем третью строчку X_3. Если это либо аксиома $\overline{M}_1$, либо имеет форму BY, где Y есть предшествующие строки (X_1 или X_2), тогда X_3 истинна в M_2 по тем же причинам, как и прежде. Если X_3 не является ни одной из них, тогда она должна быть выведена из X_1 и X_2 по правилу *modus ponens*, и так как мы уже знаем, что X_1 и X_2 истинны в M_2, отсюда следует, что X_3 истинна в M_2. (Ясно, что для любых предложений X и Y, если X и $X \supset Y$ истинны в M_2, тогда у истинно в M_2). Сейчас мы знаем, что первые три строчки доказательства истинны в M_2, и с этим знанием истинность X_4 может быть установлена той де самой аргументацией. Тогда, зная истинность первых четырех строчек, мы подобным же образом получаем истинность пятой строчки, и так далее, пока не доходим до последней строчки в доказательстве[1]. Это заключает доказательство Леммы А.

Самореферентная правильность систем $\overline{K}$, $\overline{K}_4$ и $\overline{G}$ следует из Следствия A_1 и следующей Леммы.

Лемма В. Для любой модальной системы М:

(a) Если М принадлежит к типу 1, тогда все аксиомы $\overline{K}$ истинны в М.

(b) Если М нормальна и принадлежит к типу 1, тогда все аксиомы $\overline{K}_4$ истинны в М.

(c) Если М принадлежит к типу G, тогда все аксиомы $\overline{G}$ истинны в М.

2.

Почему Лемма В верна?

Решение: (а) Предположим, что множество доказуемых предложений М замкнуто относительно *modus ponens*. Все тавтологии истинны в М (они истинны в любой модальной системе). Другими аксиомами $\overline{K}$ являются предложения формы $(BX\&B(X\supset Y))\supset BY$, или же, альтернативно, $B(X\supset Y)\supset(BX\supset BY)$; между ними нет разницы. Сказать, что $(BX\&B(X\supset Y))\supset BY$ истинно в М, значит сказать, что если X доказуемо в М и $X\supset Y$ доказуемо в М, доказуемо в М и Y. Ну, это и имеет место, так как доказуемые предложения в М замкнуты относительно *modus ponens*.

(b) Предположим, что М есть нормальная система типа 1. Тогда все аксиомы $\overline{K}$ истинны в М – по (а). Другие

[1] Этот тип аргументации известен под названием *математической индукции*.

аксиомы $\overline{K}_4$ есть предложения формы $BX \supset BBX$. Ну, сказать, что такое предложение истинно в M, значит сказать, что если BX истинно в M, то в M истинно и BBX. Другими словами, если X доказуемо в M, то доказуемо и BX. Это действительно так, поскольку M нормальна.

(c) Предположим, что M принадлежит к типу G. Тогда она определенно типа 4, поэтому, по (b), все аксиомы $\overline{K}_4$ истинны в M. Остальные аксиомы $\overline{G}$ – это предложения формы $B(BX \supset X) \supset BX$, и сказать, что они *истинны* в M, значит сказать, что если $BX \supset X$ доказуемо в M, тогда доказуемо и X. Другими словами, система M является лёбовой. Это заключает доказательство Леммы B.

Мы продолжаем предполагать, что M есть любая из систем K, K_4 или G. Тогда $\overline{M}$ есть соответственно $\overline{K}$, $\overline{K}_4$ или $\overline{G}$.

Следствие B_1. Все аксиомы $\overline{M}$ истинны в M.

Следствие B_2. Все аксиомы $\overline{M}$ истинны в M.

Решение: Так как $\overline{K}$ принадлежит типу 1, $\overline{K}_4$ нормальна и принадлежит типу 1, а $\overline{G}$ принадлежит типу G, Следствие B_1 выводится непосредственно из Леммы B. Также, поскольку K принадлежит к типу 1, K_4 нормальна и принадлежит к типу 1, а G принадлежит к типу G, мы также имеем Следствие B_2.

Из Следствия B_1 и Следствия A_1 мы имеем Теорему 1.

Теорема 1. Системы $\overline{K}$, $\overline{K}_4$ и $\overline{G}$ являются самореферентно правильными.

Следствие. Системы $\overline{K}$, $\overline{K}_4$ и $\overline{G}$ непротиворечивы и устойчивы.

Мы теперь имеет третий пример непротиворечивой и устойчивой системы типа G, а именно модальную систему $\overline{G}$ (другими двумя системами были машины из предыдущей главы). Мы вскоре увидим, что модальная система G также является самореферентно правильной (и отсюда, непротиворечивой и устойчивой).

Я надеюсь, что читатель полностью понимает абсурдность сомнений в непротиворечивости системы просто на том основании, что она не может доказать свою собственную непротиворечивость!

Системы K, K_4 и G. Для установления самореферентной правильности систем K, K_4 и G мы учтем следующее. Прежде всего, Лемма A имеет другое следствие.

Следствие A_2. Пусть M будет любой модальной системой, чьими единственными правилами вывода являются *modus ponens* и вынуждение. Тогда, если все аксиомы $\overline{M}$ истинны в M, все доказуемые предложения $\overline{M}$ истинны в M.

Из Следствия A_2 и Следствия B_2 выводится, что если M есть любая из модальных систем K, K_4 или G, тогда все истинные предложения, доказуемые в $\overline{M}$, истинны в M. Но у нас пока кое-чего нет. Остается показать, что если M является любой из трех систем K, K_4 или G, тогда любое предложение, доказуемое в M, также доказуемо и в $\overline{M}$ (факт, который утверждался без доказательства в конце главы 19). Раз это сделано, доказательство самореферентной правильности K, K_4 и G будет полным.

3.

Почему истинно, что если предложение доказуемо в M (M есть любая из K, K_4 или G), тогда оно доказуемо в $\overline{M}$?

Решение: Исследованием любой из трех систем K, K_4 или G можно видеть, что если X есть аксиома M, тогда если мы подставим любые предложения для пропозициональных переменных в X (подставляя одно и то же предложение в различные вхождения той же самой переменной), результирующее предложение есть также аксиома M – и отсюда, $\overline{M}$. Мы сейчас имеем конкретное предложение, скажем T, и для любой формулы X, пусть X' будет результатом подстановки T для *всех* пропозициональных переменных в X. Конечно, если X есть само предложение, мы берем X' в качестве X. Мы теперь замечаем следующие факты: (1) Если X есть аксиома M, тогда X' есть аксиома $\overline{M}$. (2) Для любых формул X и Y, предложение $(X \supset Y)'$ есть предложение $(X' \supset Y')$, и следовательно, для любых формул X, Y и Z, если Z выводимо из X и Y посредством *modus ponens*, тогда Z' выводимо из X' и Y' посредством *modus ponens*. (3) Для любой формулы X предложение (BX)' есть предложение BX' (то есть B, за которой следует X), и поэтому, если Y выводимо из X по правилу вынуждения, тогда Y' выводимо из X' по правилу вынуждения. Отсюда следует, что для данной последовательности $X_1, \ldots, X_n$ формул, если эта последовательность составляет доказательство в системе M, тогда последовательность $X'_1, \ldots, X'_n$ составляет доказательство в $\overline{M}$. И поэтому, если X есть некоторая формула, доказуемая в M, предложение X' тогда доказуемо в $\overline{M}$. Если, далее, X окажется предложением, тогда X = X', и отсюда, X само доказуемо в $\overline{M}$. Это доказывает, что любое предложение, доказуемое в M, доказуемо и в $\overline{M}$.

ЧАСТЬ XI

ФИНАЛ

ГЛАВА 28. *Модальные системы, Машины и Мыслители*

В данной главе мы встретимся с некоторыми очень странными Мыслителями. Для того чтобы полностью оценить их, давайте сначала обратимся к теме минимальных Мыслителей.

Минимальные Мыслители различных типов

Модальное предложение X само по себе не истинно, ни ложно; оно выражает определенное суждение, если только символу «B» дана интерпретация. Мы определили предложение как истинное в модальной системе M, если оно *истинно*, когда «B» интерпретировано как доказуемость в M. Мы будем далее считать модальное предложение истинным для Мыслителя, если оно истинно, когда «B» интерпретировано как *Мыслитель верит в*.

Сказать, что предложение *истинно* для Мыслителя, значит сказать нечто совершенно отличное от того, что Мыслитель *верит в* него. Например, сказать, что $\neg B\bot$ истинно для Мыслителя, значит сказать, что Мыслитель непротиворечив, в то время как сказать, что Мыслитель верит в $\neg B\bot$, значит сказать, что Мыслитель верит в свою собственную непротиворечивость.

Машина может быть запрограммирована так, чтобы печатать все и только те предложения, которые доказуемы в модальной системе $\bar{G}$, заданием машине следующих инструкций: (1) На некоторой стадии вы можете напечатать любую аксиому $\bar{G}$. (2) Если на некоторой стадии вы можете напечатать предложения X и $(X \supset Y)$, тогда вы можете напечатать Y. (3) Если на некоторой стадии вы напечатали X, тогда вы можете напечатать BX. (Машине можно дать дальнейшие инструкции, которые будут гарантировать, что все, что машина *может* делать, она *сделает* рано или поздно, и поэтому каждое предложение, доказуемое в $\bar{G}$, будет рано или поздно напечатано машиной). Давайте назовем такую машину $\bar{G}$-машиной.

Теперь представим Мыслителя, который непрерывно наблюдает за выходом машины. Однако он интерпретирует BX не как «X печатаемо машиной» и не как «X доказуемо в $\bar{G}$», а как «Я верю в X». (Он полагает, что машина печатает предложения о *нем*!) Его интерпретация может быть названа *эгоцентричной* интерпретацией модальных предложений.

Предположим, что Мыслитель испытывает полную уверенность в том, что машина знает, во что он верит, и поэтому всякий раз, когда машина печатает X, он немедленно верит в это (конечно, при эгоцентричной интерпретации). Его система веры будет тогда включать *все* предложения, доказуемые в $\bar{G}$. Это *не* гарантирует, что он принадлежит к типу G (он может не быть нормальным, несмотря на то что верит в свою нормальность, и он может не быть даже типа 1, несмотря на то что он верит, что принадлежит к типу 1). Конечно, если он *правильно* верит во все предложения $\bar{G}$, тогда легко видеть, что он типа G.

Но теперь предположим, что он верит в *те и только в те* предложения, которые печатаемы машиной; его система вер тогда в точности совпадает с множеством предложений, доказуемых в $\bar{G}$, и так как $\bar{G}$ принадлежит к типу G, *он* должен быть типа G. Такого Мыслителя мы назовем минимальным Мыслителем типа G. Так как система $\bar{G}$ самореференциально правильна (как мы показали в предыдущей главе), минимальный Мыслитель типа G должен быть полностью точен в своих верах. Кроме того, следует, что любой минимальный Мыслитель типа G одновременно непротиворечив и устойчив.

Мы сейчас видим, что понятие непротиворечивого устойчивого Мыслителя типа G не включает в себя логического противоречия. Мыслитель типа G не необходимо непротиворечив (на самом деле, любой *противоречивый* Мыслитель даже типа 1 принадлежит также к типу G, так как он верит во все что угодно!), но минимальный Мыслитель типа G одновременно непротиворечив и устойчив.

Теперь рассмотрим Мыслителя, который *принадлежит* к типу G и который непрерывно наблюдает за выходом машины (интерпретируя при этом все модальные предложения эгоцентрично). Будет ли он необходимо *верить* во все эти предложения? Ну, легко проверить, что он верит во все *аксиомы* $\overline{G}$. (На самом деле, в Главе 11 мы показали, что любой Мыслитель типа 4 знает, что он принадлежит к типу 4; отсюда, он верит во все аксиомы $\overline{K}_4$. Мыслитель типа G также верит в свою скромность, что означает, что он поверит во все предложения формы $B(BX \supset X) \supset BX$, и поэтому он верит во все аксиомы $\overline{G}$.) И так как машина печатает предложения, не являющиеся аксиомами, только с помощью правил *modus ponens* и вынуждения, и веры Мыслителя замкнуты относительно *modus ponens*, и он нормален, он будет успешно верить в каждое предложение, которое будет напечатано. (Если кто-нибудь прервет процесс и спросит мнение Мыслителя о машине, Мыслитель ответит: «Это машина воистину восхитительна. Все, что она печатает обо мне, до сих пор оказывалось истинным!»)

Мы сейчас видим, что Мыслитель типа G в самом деле верит во все предложения, доказуемые в $\overline{G}$.

Предположим, далее, что в модальное предложение X верят все Мыслители типа G; следует ли отсюда, что X на самом деле доказуемо в $\overline{G}$? Ответ на этот вопрос утвердителен, так как если в X верят *все* Мыслители типа G, тогда в него должен верить *минимальный* Мыслитель типа G, и, отсюда, X должно быть доказуемо в $\overline{G}$. И поэтому мы сейчас видим, что предложение доказуемо в $\overline{G}$, *если и только если*, в него верят все Мыслители типа G. Другими словами, пусть имеется некоторый минимальный Мыслитель типа G, тогда он верит в те и только те предложения, в которые верят *все* Мыслители типа G.

Конечно, когда два различных Мыслителя рассматривают одно и то же модальное предложение, они интерпретируют его по-разному – каждый интерпретирует «B» в соотнесении со своими *собственными* верами (точно так же слово «я» имеет различный референт, когда оно используется различными людьми). И поэтому, когда мы говорим о модальном предложении, в которое верят все Мыслители типа G, мы имеем в виду, что каждый верит в него в соответствии с собственной эгоцентричной интерпретацией.

Конечно, все сказанное о модальной системе $\overline{G}$ и Мыслителях типа G справедливо также и для модальной системы $\overline{K}_4$ и Мыслителей типа 4: Предложение доказуемо в $\overline{K}_4$, если и только если в него верят все Мыслители типа 4. Подобным же образом, предложение доказуемо в $\overline{K}$, если и только если, в него верят все Мыслители типа 3.

Больше о модальных системах и Мыслителях

Материал, изложенный в оставшейся части этой главы, не является необходимым для понимания следующих двух глав, но представляет независимый интерес.

1.

Предположим, что система вер Мыслителя замкнута относительно *modus ponens* и что для любой *аксиомы* X системы $\overline{K}_4$. Мыслитель верит в X, а также верит, что он верит в X. Будет ли он необходимо верить во все предложения, в которые верят все Мыслители типа 4? (Вспомним, что может и не быть нормальным!) Ответ дается в следующей Проблеме 2.

2.

Теперь подставим $\overline{G}$ вместо K_4. Будет ли Мыслитель, описанный выше, необходимо верить во все предложения, в которые верят все Мыслители типа G?

Ответ на эту проблему дается хорошо известной теоремой о модальных системах K_4 и G (которая также применима и к системам $\overline{K}_4$, и $\overline{G}$), которую мы готовы установить и доказательство которой мы набросаем.

Пусть M будет модальной системой, чьими единственными правилами вывода являются *modus ponens* и вынуждение. Пусть M' будет модальная система, аксиомы которой совпадают с аксиомами для системы M, вместе со всеми формулами BX, где X есть аксиома системы M, и чье *единственное* правило вывода есть *modus ponens*. Ясно, что все доказуемое в системе M', доказуемо в системе M (потому что для любой аксиомы X из M, BX также доказуемо в M, и поэтому все аксиомы M' доказуемы в M), но в общем, неверно, что все доказуемое в M также доказуемо в M'. Однако мы имеем следующий интересный результат:

Теорема 1. Если все аксиомы K_4 доказуемы в M, тогда верно, что все доказуемое в M доказуемо в M' (и, отсюда, системы M и M' доказывают в точности одни и те же формулы).

Может ли читатель понять, как доказывать Теорему 1?

(Подсказка: Сначала покажем, что если X доказуемо в M, тогда BX доказуемо в M'. Это делается демонстрацией того, что для любого доказательства $X_1, \ldots, X_n$ в M, все формулы $BX_1, \ldots, BX_n$ последовательно доказуемы в M'.)

Более детальное доказательство. Так как *modus ponens* есть правило в M' и все тавтологии находятся среди аксиом M', тогда M', конечно же, принадлежит к типу 1. Также для M' справедливы следующие три условия:

(1) Если BX и $B(X \supset Y)$ доказуемы в M', тогда доказуемо и BY – поскольку $B(X \supset Y) \supset (BX \supset BY)$ есть аксиома M', а M' типа 1.

(2) Если BX доказуемо в M', тогда доказуемо и BBX – потому что $BX \supset BBX$ есть аксиома M', а M' принадлежит к типу 1.

(3) Если X есть аксиома M, тогда BX доказуемо в M' – потому что даже аксиома M'.

Теперь предположим, что последовательность $X_1, \ldots, X_n$ формул составляет доказательство в M. Каждая строка доказательства получается либо как результат двух предыдущих строк по правилу *modus ponens*, или же из предшествующей строки по правилу вынуждения, или же сама является аксиомой M. Используя факты (1), (2) и (3) выше, легко вывести, что BX_1 должно быть доказуемо в M', затем, что BX_2 доказуемо в M', затем BX_3 доказуемо в M', и так далее, вплоть до BX_n. Мы оставляет проверку этого читателю.

Сейчас мы знаем, что если X доказуемо в M, тогда BX доказуемо в M'. Следовательно, если X доказуемо в M', тогда BX доказуемо в M' (потому что если X доказуемо в M', тогда оно также доказуемо в M). И поэтому мы видим, что M' нормально (хотя правило вынуждения исходно не задано в M'). Тогда, если дано некоторое доказательство $X_1, \ldots, X_n$ в M, легко видеть, что каждая из строк $X_1, \ldots, X_n$ сама может быть успешно доказана в M'. (Мы оставляем проверку этого читателю.)

Следствие. Доказуемые формулы K_4 и K_4' совпадают. Доказуемые формулы G и G' совпадают.

Следующая теорема и ее следствие могут быть легко доказаны аналогичным образом.

Теорема $\bar{1}$. Для любой сентенциональной модальной системы $\bar{M}$, в которой доказуемы все аксиомы системы $\bar{K}_4$ и в которой правила *modus ponens* и вынуждения являются единственными, доказуемые предложения M' совпадают с доказуемыми предложениями $\bar{M}$'.

Следствие. Доказуемые предложения $\bar{K}_4$ совпадают с доказуемыми предложениями $\bar{K}_4$'. Доказуемые предложения $\bar{G}$ совпадают с доказуемыми предложениями $\bar{G}$'.

Конечно же, следствие выше дает утвердительный ответ на Проблемы 1 и 2.

Мы видим, благодаря Следствию Теоремы 1, что модальные системы K_4 и G могут быть аксиоматизированы альтернативным образом, используя системы, чьим единственным правилом вывода является modus ponens.

ГЛАВА 29. *Некоторые странные Мыслители!*

Мыслители, которые почти типа G

Мы будем говорить, что Мыслитель *почти* типа G, если он верит во все предложения, в которые верят все Мыслители типа G (или, что то же самое, он верит во все предложения, доказуемые в модальной системе $\bar{G}$), и если его веры замкнуты относительно *modus ponens*. Его удерживает от принадлежности к типу G то, что он может не быть нормальным.

Как мы покажем, Мыслитель, который почти типа G, в отличие от Мыслителей типа G, *может* поверить, что он непротиворечив, не утрачивая собственной непротиворечивости. Но тогда он должен страдать от другого дефекта – он не может быть нормальным!

Мы рассмотрим эту тему более подробно.

1.

Пусть имеется непротиворечивый Мыслитель почти типа G, который верит в свою собственную непротиворечивость. Найти такое предложение p, что Мыслитель верит в p, но никогда не знает, что он верит в p!

Решение: Одно такое предложение p таково: Мыслитель непротиворечив!

Нам дано, что он верит в $\neg B\bot$, и нам нужно показать, что, если он непротиворечив, он не может верить в $B\neg B\bot$. Ну, так как он верит во все предложения, доказуемые в G, он определенно верит во все тавтологии, и так как его веры замкнуты относительно *modus ponens*, он определенно принадлежит к типу 1. Он верит в $\neg B\bot$, поэтому он верит в $B(B\bot\supset\bot)$. Если он верит в $B\neg B\bot$, он должен поверить в $B(B\bot\supset\bot)$. Однако он таки верит в $B(B\bot\supset\bot)\supset B\bot$ (потому что он верит во все предложения, доказуемые в G). И поэтому он поверит тогда $B(B\bot\supset\bot$ и в $B(B\bot\supset\bot)\supset B\bot$, и отсюда, он верит в $B\bot$. Но так как он верит в$\neg B\bot$, он будет противоречив.

Это доказывает, что если он верит в $B\neg B\bot$, он будет противоречив. Но нам дано, что он непротиворечив, отсюда, он никогда не поверит в $B\neg B\bot$ (несмотря на то что $B\neg B\bot$ истинно).

2.

Любой Мыслитель, не являющийся нормальным, не сможет поверить в по крайней мере одно истинное предложение, потому что имеется предложение p такое, что он верит в p, но не может поверить в Bp, и все же Bp истинно (так как он верит в p). Следовательно, он не сможет поверить в истинное предложение Bp.

Из последней Проблемы тогда получается вот что. Пусть дан некоторый непротиворечивый Мыслитель, который почти типа G, который верит в свою собственную непротиворечивость. Тогда должно быть, по крайней мере, одно истинное предложение, в которое он не может поверить. Более озадачивающим фактом является то, что должно быть по крайней мере одно *ложное* предложение, в которое он таки верит.

Какое это ложное предложение?

Решение: $B\neg B\bot$ истинно, но так как он не верит в него, $BB\neg B\bot$ ложно, отсюда, $B\neg B\bot\supset BB\neg B\bot$ ложно. Но он должен верить в это предложение (потому что он формы $BX\supset BBX$, где $X=\neg B\bot$), отсюда, это аксиома системы G. И поэтому он верит в ложное предложение $B\neg B\bot\supset BB\neg B\bot$. (Он верит неправильно: «Если я должен был бы верить, что я непротиворечив, тогда я должен был бы верить, что я верю в свою непротиворечивость». Эта вера неверна, так как на самом деле он *все-таки верит*, что он непротиворечив, но не верит – и никогда не поверит – что он верит в свою собственную непротиворечивость.)

Кстати говоря, по той же аргументации, любой Мыслитель, не являющийся нормальным, который верит во все предложения, доказуемые в K_4, должен иметь по крайней мере одну ложную веру. Имеется некоторое p такое, что он верит в p, но не верит в Bp, поэтому $Bp\supset BBp$ ложно (для такого Мыслителя), и все же это аксиома K_4, и следовательно, Мыслитель верит в него. Это ответ на Упражнение 2 ниже.

Упражнение 1. Установить, является ли следующее истинным или ложным: Каждый Мыслитель типа 1, не являющийся нормальным, непротиворечив.

Решение: Оно истинно! Если бы он был противоречивым и принадлежал типу 1, он верил бы во *все* предложения, отсюда, не было бы предложения p такого, что он верил бы в p и не верил бы в Bp (потому что он верит в p и Bp, отсюда, он верит во все, что угодно). Следовательно, каждый противоречивый Мыслитель типа 1 должен быть нормальным – или же, другими словами, каждый Мыслитель, не являющийся нормальным и принадлежащим к типу 1, непротиворечив.

Упражнение 2. Установить, является ли следующее истинным или ложным: Каждый Мыслитель типа 1, не являющийся нормальным, который верит во все аксиомы [K_4], должен верить в, по крайней мере, одно ложное предложение.

Мыслители, принадлежащие типу G*

Как мы увидим, Мыслитель, который почти типа G, может не только верить в свою собственную непротиворечивость, не будучи при этом необходимо противоречивым; он может даже верить в свою собственную *точность*, не будучи при этом необходимо противоречивым.

Под Мыслителем типа G* мы будем иметь в виду Мыслителя, который почти типа G и который верит во все предложения формы $BX \supset X$ (он верит в свою собственную точность). Другими словами, Мыслитель типа G* есть Мыслитель типа 1, который верит во все предложения, доказуемые в G, и верит во все предложения формы $BX \supset X$.

Такой Мыслитель должен, конечно, верить в предложение $B\bot \supset \bot$, и так как он почти типа G, тогда то, что мы доказали в решении Проблем 1 и 2, справедливо и для него. И поэтому мы формулируем Теорему 1.

Теорема 1. Для любого непротиворечивого Мыслителя типа G*:

(a) Он верит в свою собственную непротиворечивость, но никогда не сможет узнать, что он верит в то, что непротиворечив!

(b) Он также верит в ложное предложение $B\neg B\bot\supset\supset BB\neg B\bot$ (то есть он неправильно верит в «если я когда-либо поверю в свою собственную непротиворечивость, я поверю, что я верю в свою собственную непротиворечивость».)

Мы напоминаем читателю, что предложение $B\neg B\bot\supset\supset BB\neg B\bot$ ложно для непротиворечивого Мыслителя типа G*, так как $B\neg B\bot$ истинно (он верит в свою собственную непротиворечивость), но $BB\neg B\bot$ ложно (он не верит, что верит в свою собственную непротиворечивость).

Из Теоремы 1, конечно же, следует, что *любой* Мыслитель типа G* должен имеет, по крайней мере, одну ложную веру, потому что, если он непротиворечив, тогда он имеет ее (по Теореме 1), а если он противоречив, он определенно имеет ее!

Минимальные Мыслители типа G^*. Под модальной системой G* имеется в виду система, чьи аксиомы являются доказуемыми формулами G вместе со всеми формулами формы $BX\supset X$ и чьим единственным правилом вывода является *modus ponens*. Пусть $\bar{C}^*$ будет системой G^*, чьи аксиомы ограничены предложениями – то есть аксиомы $\bar{C}^*$ являются предложениями, доказуемыми в $\bar{C}$, плюс *предложения* формы $BX\supset X$. Единственным правилом вывода является *modus ponens*. (Мы можем легко показать с помощью аргументации, подобной той, которая использовалась в Главе 27, что доказуемые предложения $\bar{C}^*$ совпадают с доказуемыми *предложениями* G*.) Под *минимальным* Мыслителем типа G* мы имеем в виду Мыслителя, который верит в те и только в те предложения, которые доказуемы в $\bar{G}^*$. Легко показать, что все Мыслители типа G* должны верить во все предложения, доказуемые в $\bar{G}^*$ («верить» здесь, конечно, подпадает под эгоцентричную интерпретацию). Следовательно, Мыслитель есть минимальный Мыслитель типа G*, если и только если он верит в те и только в те предложения, в которые верят все Мыслители типа G*.

Так как каждый Мыслитель типа G* имеет, по крайней мере, одну ложную веру, такую веру имеет и минимальный Мыслитель типа G*. Отсюда следует, что имеется по крайней мере одно предложение, доказуемое в $\bar{G}^*$, которое ложно в $\bar{G}^*$ (ложно, когда «B» интерпретируется как доказуемость в $\bar{G}^*$). И поэтому мы имеем Теорему 2.

Теорема 2. Модальная система $\bar{G}^*$ *не* является самореферентно правильной.

В свете Теоремы 2, читатель может удивиться, как вообще можно использовать модальную систему $\bar{G}^*$. Но дело в том, что наличие доказуемого предложения в системе $\bar{G}^*$, которое ложно в $\bar{G}^*$, не означает, что не существует другой интерпретации «B», при которой все доказуемые предложения $\bar{G}^*$ истинны. Есть ли такая интерпретация? Да, есть, и весьма важная.

Теорема 3. Каждое предложение, доказуемое в $\overline{G}^*$, истинно в модальной системе $\overline{G}$.

Это означает, что каждое предложение, доказуемое в $\overline{G}^*$, истинно, если и только если, «B» *интерпретируется как доказуемость в* $\overline{G}$, *а не как доказуемость в* $\overline{G}^*$.

3.

Почему Теорема 3 верна?

Решение: Мы доказали в последней главе, что G самореферентно правильна, отсюда:

(1) Все предложения, доказуемые в $\overline{G}$, истинны в $\overline{G}$.
Также:
(2) Все предложения формы $BX \supset X$ истинны в $\overline{G}$.

Причина для (2) состоит в том, что если BX истинна в $\overline{G}$, тогда X доказуемо в $\overline{G}$ (это и есть то, что значит для BX быть истинным в $\overline{G}$), и X должно быть, следовательно, истинным в $\overline{G}$ (так как $\overline{G}$ самореферентно правильно). И поэтому, если BX истинно в $\overline{G}$, истинным является и X – что означает, что $BX \supset X$ истинно в $\overline{G}$.

Благодаря (1) и (2) каждая аксиома $\overline{G}^*$ истинна в $\overline{G}$. Так как единственное правило вывода в $\overline{G}^*$ – это *modus ponens* и так как множество предложений, истинных в $\overline{G}$, замкнуто относительно *modus ponens* (если X и $X \supset Y$ истинно в $\overline{G}$, ясно, что Y истинно в $\overline{G}$), отсюда следует, что каждое предложение, доказуемое в $\bar{G}^*$, должно быть истинно в $\bar{G}$. Таким образом, система корректна для $\bar{G}$.

Теорема 4 есть простое следствие Теоремы 3.

Теорема 4. Система $\bar{G}^*$ непротиворечива.

4.

Почему Теорема 4 есть следствие Теоремы 3?

Решение: Так как $\bar{G}^*$ правильна для $\bar{G}$, тогда если бы $\perp$ было доказуемо в $\bar{G}^*$, оно было был истинно в $\bar{G}$, что абсурдно. Следовательно, $\perp$ не доказуемо в $\bar{G}^*$, и поэтому $\bar{G}^*$ непротиворечиво (хотя не является самореферентно правильной).

Теорема 4, конечно, влечет, что любой *минимальный* Мыслитель типа G* непротиворечив. И поскольку минимальный Мыслитель типа G* непротиворечив, он верит в свою собственную непротиворечивость, но не сможет никогда поверить, что он верит в свою собственную непротиворечивость (по Теореме 2). В терминах модальной системы $\bar{G}^*$ этот результат может быть установлен альтернативно: она непротиворечива, она может доказать свою собственную непротиворечивость, но никогда не сможет доказать, что может доказать свою собственную непротиворечивость! А также модальная система $\bar{G}^*$ не является нормальной.

Полнота $\bar{G}^*$ для $\bar{G}$. Мы сейчас установим дальнейший результат, доказательство которого, к несчастью, выходит за пределы этой книги.

Пусть даны две модальные системы M_1 и M_2. Мы определяем M_1 как правильную для M_2, если каждое предложение, доказуемое в M_1, истинно для M_2. Давайте говорить, что M_1 полна для M_2, если каждое предложение, которое истинно для M_2, доказуемо в M_1.

Теорема 3 говорит, что модальная система $\overline{G}^*$ *правильна* для модальной системы $\overline{G}$. Ну, она также *полна* для $\overline{G}$ – каждое предложение, *истинное* для $\overline{G}$, *доказуемо* в $\overline{G}^*$. И поэтому доказуемые предложения $\overline{G}^*$ есть в точности те предложения, которые истинны для $\overline{G}$. Таким образом, предложение доказуемо в $\overline{G}^*$, если и только если, оно истинно для всех Мыслителей типа G.

Мыслители типа Q
(мнительные Мыслители)

Под *мнительным* Мыслителем – или Мыслителем типа Q – мы будем иметь в виду Мыслителя типа G, который верит в то, что он *противоречив*. Может ли мнительный Мыслитель быть непротиворечивым? Мы вскоре увидим, что может! Конечно же, каждый мнительный Мыслитель нормален.

Под модальной системой $\overline{Q}$ мы будем иметь в виду модальную систему $\overline{G}$ с добавленным в виде аксиомы предложением $B\bot$. Под *минимальным* Мыслителем типа Q мы имеем в виду Мыслителя, который верит в те и только те предложения, которые доказуемы в модальной системе [Q], или, что то же самое, Мыслителя, который верит во все те и только те предложения, в которые верят все Мыслители типа Q.

Теорема 5. Модальная система $\bar{Q}$ не является самореферентно правильной, но она непротиворечива.

5.

Почему Теорема 5 истинна?

Решение: Система $\bar{Q}$ принадлежит к типу G, и следовательно, согласно (с) Леммы В из Главы 27, все аксиомы $\bar{G}$ истинны для $\bar{Q}$.

Теперь давайте на момент предположим, что предложение $B\perp$ истинно для $\bar{Q}$. Мы тогда получим следующее противоречие: Если $B\perp$ истинно для $\bar{Q}$, тогда, поскольку все другие аксиомы для $\bar{Q}$ (то есть, аксиомы для $\bar{G}$) истинны для $\bar{Q}$, мы имели бы истинность *всех* аксиом $\bar{Q}$ для $\bar{Q}$. Тогда, по Следствию A_1 Леммы А из Главы 27, система $\bar{Q}$ самореферентно правильна. И поэтому, если $B\perp$ истинно для $\bar{Q}$, тогда $\bar{Q}$ самореферентно правильна. С другой стороны, сказать, что $B\perp$ истинно для $\bar{Q}$, значит сказать, что $\perp$ доказуемо в $\bar{Q}$, и так как $\perp$ явно ложна для $\bar{Q}$, это означало бы, что $\bar{Q}$ *не* является самореферентно правильной. Следовательно, противоречиво утверждать, что $B\perp$ истинно для $\bar{Q}$. Отсюда, $B\perp$ ложно для $\bar{Q}$, что означает, что $\perp$ *не* доказуемо в $\bar{Q}$, и поэтому $\overline{Q}$ должно быть непротиворечивым! Но $B\bot$ также есть аксиома $\overline{Q}$, отсюда, оно, конечно, доказуемо в $\overline{Q}$, и так как оно ложно для $\overline{Q}$, $\overline{Q}$ не является самореферентно правильной. И поэтому мы видим, что $\overline{Q}$ непротиворечиво, но не самореферентно правильна.

Из Теоремы 5, конечно, следует, что любой минимальный Мыслитель типа Q непротиворечив, хотя и не верит, что является таковым.

Сравнение. Поучительно и занимательно сравнить минимальных Мыслителей, принадлежащих типам G, G* и Q.

(1) Минимальный Мыслитель типа G непротиворечив, но никогда не может узнать это.
(2) Минимальный Мыслитель типа G* непротиворечив, верит в свою собственную непротиворечивость, но никогда не сможет узнать, что он верит в собственную непротиворечивость.
(3) Минимальный Мыслитель типа Q верит, что он противоречив, но он неправ – на самом деле он непротиворечив.

ГЛАВА 30. *В ретроспективе*

Мы начали наше исследование с интроспективных Мыслителей и углубились в лабиринты модальной логики. Давайте подведем некоторые основные итоги, полученные в ходе нашего путешествия.

1. Точная геделева система типа 1 не может доказать своей собственной точности, то есть не может доказать всех предложений формы $BX \supset X$.

2. Любая геделева система типа 1, которая может доказать свою собственную точность, не является точной, но является странной – должно быть предложение p такое, что p и $\neg Bp$ оба доказуемы.

3. Любая геделева система типа 1*, которая может доказать свою не-странность, является странной.

4. (Первая Теорема о Неполноте Геделя). Любая нормальная, устойчивая, непротиворечивая геделева система типа 1 должна быть неполной. Более детально, если S есть нормальная система типа 1, а p есть предложение такое, что $p \equiv \neg Bp$ доказуемо в S, тогда:

   (a) Если S непротиворечива, p не доказуемо в S.

   (b) Если S непротиворечива и устойчива, тогда $\neg p$ также не доказуемо в S.

5. (Вторая Теорема о Неполноте Геделя.) Никакая геделева система типа 4 не может доказать своей собственной непротиворечивости.

6. Геделева система типа 4 может даже доказать, что если она непротиворечива, тогда она не может доказать

своей собственной непротиворечивости, то есть может доказать предложение $\neg B\bot \supset \neg B(\neg B\bot)$.

7. (Лёб.) Если S есть рефлексивная система типа 4, тогда для любого предложения p системы, если $Bp \supset p$ доказуемо в системе, тогда доказуемо и p.

8. Система типа 4 рефлексивна, если и только если, она типа G.

9. Система типа 4 является лёбовской, если и только если, она типа G.

10. (Крипке, Де Йонг, Самбин.) Любая система типа 3, в которой все предложения формы $B(BX \supset X) \supset BX$ доказуемы, должна быть типа 4.

11. Непротиворечивая система типа G не может доказать никаких предложений формы $\neg BX$, в частности она не может доказать собственной непротиворечивости.

12. Непротиворечивая и стабильная система типа G не может доказать ни своей собственной непротиворечивости, ни своей собственной противоречивости.

13. (Семантическая Теорема Адекватности.) Для любой модальной формулы X, если X доказуема в K, тогда она справедлива во всех моделях Крипке; если она доказуема в K_4, она справедлива во всех транзитивных моделях; и если она доказуема в G, тогда она справедлива во всех транзитивных терминальных моделях.

14. Существуют таки непротиворечивые устойчивые системы типа G, например машины Фергюссона и Крейга, и модальная система $\bar{G}$. Эти системы, несмотря на свою непротиворечивость, не могут доказать ее.

15. Модальные системы $\bar{K}$, $\bar{K}_4$, $\bar{G}$ являются не только непротиворечивыми и стабильными, но и самореферентно правильными. То же самое для систем K, K_4 и G.

16. Ни одна из систем $\bar{G}^*$ и $\bar{Q}$ не является референтно правильной, но обе они непротиворечивы. Система $\bar{Q}$ нормальна, а система $\bar{Q}$ не нормальна.

17. (a) Минимальный Мыслитель типа G непротиворечив, но никогда не узнает об этом.

(b) Минимальный Мыслитель типа G* непротиворечив и верит, что он непротиворечив, но он никогда не узнает, что верит в свою собственную непротиворечивость.

(c) Минимальный мыслитель типа Q верит, что он противоречив, но на самом деле он непротиворечив.

Судя по всему, тут пора и остановиться. Есть много восхитительных вещей о модальной системе G. Самый лучший источник сведений по этому поводу – это книга Булоса *Unprovability of Consistency*[1], которую я рекомендую всем после прочтения моей книги. Чтобы разжечь аппетит читателя, позвольте мне привести пример прекрасного результата – теорему о неподвижной точке, доказательство которой можно найти в упомянутой книге.

Рассмотрим модальную формулу с одной пропозициональной переменной, скажем, буквой p. Давайте обозначим такую формулу через A(p). Для любого модального предло-

[1] Как уже упоминалось выше, Дж. Булос написал практически новую книгу по этой теме, – The Logic of Provability.

жения S, под A(S) имеется в виду результат подстановки S вместо p в A(p). Например, если A(p) есть формула $p \supset Bp$, тогда A(S) есть предложение $BS \supset S$. Предложение S называется неподвижной точкой A(p), если предложение $S \equiv A(S)$ доказуемо в системе G. В проблеме 4 Главы 19 нужно было найти предложение p такое, что любой Мыслитель типа G верит в $p \equiv \neg Bp$. Таким предложением оказалось $\neg B\bot$. Таким образом, каждый Мыслитель типа G будет верить в $\neg B\bot \equiv \equiv \neg B \neg B\bot$, и поэтому это предложение доказуемо в G. Это означает, что $\neg B\bot$ есть неподвижная точка формулы $\neg Bp$. Формула $B\neg p$ также имеет неподвижную точку, а именно $B\bot$, как мы обнаружили в Проблеме 5 Главы 19. В качестве упражнения можно показать, что формула $Bp \supset B\bot$ имеет неподвижную точку, а именно $BB\bot \supset B\bot$.

Не каждая формула A(p) имеет неподвижную точку. Например, формула $\neg p$ не имеет ее (в противном случае система G была бы противоречивой, а это, как мы знаем, не так). Формула A(p) называется *модализированной* в p, если каждое вхождение p в A(p) лежит в такой части A(p), которая имеет форму BX, где X есть формула. (Например, $Bp \supset BBp$ модализирована в p, $Bp \supset p$ нет, но $B(Bp \supset p)$ модализирована.) Логики Клаудио Бернарди и К. Сморински независимо друг от друга доказали, что любая формула A(p), которая модализирована, имеет-таки неподвижную точку S – больше того, формула $B(p \equiv A(p)) \supset B(p \equiv \Phi)$ доказуема в G. Этот результат известен как Теорема о неподвижной точке Бернарди–Сморински.

Неподвижные точки являются замечательными вещами. Благодаря самореферентной правильности системы G любая неподвижная точка S формулы A(p) не только доказуема в G, если и только если, A(S) доказуема, но также и истинна (для G), если и только если, A(S) истинна, потому что доказуемость $S \equiv A(S)$ влечет ее истинность. Давайте говорить, что формула A(p) *применима* к предложению S, если A(S) истинна (для G). Неподвижная точка формулы может тогда полагаться предложением, которое утверждает, что формула применима к самому предложению.

Более обще, рассмотрим формулу A(p,q), не имеющую пропозициональных переменных, кроме p и q. Для любой формулы X, под A(X, q) подразумевается результат подстановки X вместо p в A(p,q). Под неподвижной точкой формулы A(p,q) подразумевается формула непротиворечивая, не имеющая других переменных, помимо q, такая, что формула $H \equiv A(H,q)$ доказуема в G. Логики Д. де Йонг и Джиованни Самбин доказали, что если A(p,q) модализирована в p (но не необходимо в q), тогда A(p, q) имеет неподвижную точку H; больше того, формула $B(p \equiv A(p,q)) \supset \supset B(p \equiv H)$ доказуема в G.

Примеры этого уже известны из Главы 19. Bq есть неподвижная точка $B(p \supset q)$ (Проблема 6), а $Bq \supset q$ есть неподвижная точка формулы $Bp \supset q$ (Проблема 7). В самом деле, рефлексивность здесь эквивалентна наличию неподвижной точки для $Bp \supset q$. Примечательным здесь является то, что для модальной системы типа 4 существование неподвижной точки для одной формулы $Bp \supset q$ достаточно, чтобы гарантировать наличие неподвижных точек для *всех* формул A(p,q), которые модализированы в p. Доказательство этого можно найти в книге Булоса.

Заключительные замечания

Я надеюсь, что я дал читателю некоторое ощущение того, почему модальные системы интересны. Мы видели, что они могут быть интерпретированы как внутренне (самореферентно) и внешне (в применении к доказуемости в других математических системах), так и для анализа процесса мышления – естественных разумных существ (некоторых людей и других животных) и механизмов искусственного интеллекта (таких как компьютеры). Какое применение это может иметь в области психологии – вопрос, достойный дальнейших исследований.

Счастливым поворотом судьбы является то обстоятельство, что область модальной логики, которая исторически возникла в результате чисто философских интересов, оказалась столь важной сегодня в теории доказательства и компьютерных исследованиях. И все это благодаря теоремам Геделя и Лёба и работам тех, кто стал рассматривать теорию доказательства с теоретико-модальной точки зрения. И сейчас даже те философы, которые в прошлом имели весьма смутное представление о значимости модальной логики, вынуждены признать ее математическую важность.

Давняя философская оппозиция модальной логики зиждилась, вообще говоря, в трех совершенно различных (и несовместимых) верах. Во-первых, были такие философы, кто верил, что все истинное есть необходимо истинное, и отсюда, нет никакого различия между истиной и необходимой истиной. Во-вторых, были те, кто полагал, что ничего не является необходимо истинным, и отсюда, для любого предложения p, предложение Np (p необходимо истинно) просто ложно! И, в-третьих, были и такие философы, которые утверждали, что слова «необходимо истинный» не несут вообще никакого значения. И поэтому каждый из этих философских типов отвергал модальную логику, исходя из изложенных выше своих соображений. На самом деле, одному широко известному философу приписывается изречение, что современная модальная логика впала в грех. На что Дж. Булос ответил так: «If modern modal logic was conceived in sin, then it has been redeemed through Godliness»[1].

[1] «Мы вынуждены были оставить это изречение Булоса в английском оригинале ввиду непереводимой игры слов, основанной на сходстве слов God (Бог) и Godel (Гедель)».

Оглавление

Аннотированный список книг издательства «Канон+»
РООИ «Реабилитация» вы можете найти на сайте
iph.ras.ru/kanon или http://journal.iph.ras.ru/verlag.html
Заказать книги можно, отправив заявку по электронному адресу:
kanonplus@mail.ru

Научное издание

Раймонд СМАЛЛИАН

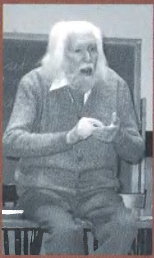
ВОВЕКИ НЕРАЗРЕШИМОЕ
Путь к Геделю через занимательные загадки

Директор – *Божко Ю. В.*
Ответственный за выпуск – *Божко Ю. В.*
Компьютерная верстка – *Липницкая Е. Е.*
Корректор – *Колупаева Л. П.*

Подписано в печать с готовых диапозитивов 03.09.2012.
Формат 84×108 1/32. Бумага офсетная. Печать офсетная.
Усл. печ. л. 15,96. Уч.-изд. л. 9,9. Тираж 1000 экз. Заказ 2662.

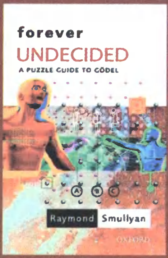
Издательство «Канон+» РООИ «Реабилитация».
111672, Москва, ул. Городецкая, д. 8, корп. 3, кв. 28.
Тел./факс 702-04-57.
E-mail: kanonplus@mail.ru
Сайт: iph.ras.ru/kanon или http://journal.iph.ras.ru/verlag.html

Республиканское унитарное предприятие
«Издательство «Белорусский Дом печати».
ЛП № 02330/0494179 от 03.04.2009.
Пр. Независимости, 79, 220013, Минск.



Рэймонд Смаллиан –
родился 25 мая 1919 в Far Rockaway, Лонг-Айленд, Нью-Йорк, США
Первым опытом работы Рэймонда стала должность сценического мага. Затем он получил степень бакалавра в Чикаго в 1955 году, а затем в 1959 году в Принстоне степень доктора наук. Занимал должность профессора философии в «Сити-колледж» в Нью-Йорке.

Рэймонд Смаллиан воплощает в одном лице единственное в своем роде собрание различных профессий: философ, логик, математик, музыкант, фокусник, юморист, писатель и составитель великолепных задач-головоломок. Искусный писатель и великолепный юморист, Смаллиан любит облекать свои задачи в литературную форму, нередко пародирующую какие-нибудь известные произведения. Делает он это настолько хорошо, что его книги, изобилующие всякого рода парадоксами, курьезами и задачами, с удовольствием читают и те, кто даже не пытается решать задачи.
Рэймонд Смаллиан является одним из многих выдающихся логиков, которые учились под руководством Алонсо Черча. Смаллиан – автор большого количества книг по занимательной математике и занимательной логике.



Эта книга представляет собой введение в теоремы Геделя посредством логических занимательных проблем с применением математической логики. Аргументация Геделя перенесена из формальной области математических систем в область идей, более доступных обычному читателю. Основной упор сделан на системы вер и их соотношению с математикой. Это приводит к семантике возможных миров, которая играет существенную роль в компьютерных исследованиях и искусственном интеллекте.

издательство
КАНОН-ПЛЮС

ISBN 978-5-88373-081-7
9 785883 730817