

В.В. Горбатов

ЧТО ПРИСНИЛОСЬ «СПЯЩЕЙ КРАСАВИЦЕ»?¹

Аннотация: В данной статье исследуются логико-онтологические аспекты рациональной оценки субъективных вероятностей и принятия решений примере парадокса «Спящей красавицы». Показывается несводимость онтологии «от первого лица» к онтологии «от третьего лица» – в частности, несводимость пропозициональных установок *de se* к установкам *de re* и *de dicto*. Естественным следствием этого является отказ от наивного референциализма в трактовке субъективных вероятностей в пользу применения более комплексных и тонких инструментов анализа – таких, например, как чалмерсовский модальный рационализм и лежащая в его основе эпистемическая двумерная семантика.

Abstract: This paper deals with "Sleeping Beauty" paradox and logico-ontological aspects of rational credences and decision-making in general. We maintain the view that "de se" propositional attitudes can not be reduced to "de re and "de dicto". The natural consequence of this is the rejection of naive referentialism in the interpretation of credences in favor of epistemic 2D semantics.

Ключевые слова: апостериорная вероятность, байесовская вероятность, проблема «спящей красавицы», модальности *de se* / *de dicto*, двумерная семантика

Keywords: posterior credence, Bayesianism, "Sleeping Beauty" problem, *de se* / *de dicto*, 2D semantics

1. Онтологизация понятия вероятности

В XX веке понятие вероятности перестало быть чисто эпистемологическим и приобрело отчётливый метафизический оттенок. Связанные с ним понятия риска, индетерминизма, случайности, шанса превратились в онтологические категории. Неточные, неопределённые и нечёткие знания получили полноправный бытий-

¹ В данной работе использованы результаты, полученные в ходе выполнения проекта 13-05-0032 «Метафизический реализм: трансцендентное в современной философии» при поддержке Программы Научный фонд НИУ ВШЭ в 2014 г.

ный статус в рамках вновь возникших вероятностных онтологий разного толка и это, конечно же, стало колоссальным шагом вперед в понимании того, какую роль играют вероятностные оценки в нашей жизни.

Вместе с тем, не следует путать *онтологизацию* вероятностей с их *объективизацией*. Если рассуждать в семантических терминах, перед нами два совершенно разных подхода к приписыванию значений пробабилистическим высказываниям. Пусть « $P(A) = n$ » – высказывание о том, что вероятность события A равняется величине n , и пусть это высказывание утверждается некоторым индивидом S . Как мы должны строить семантику для подобных утверждений? Этот вопрос разбивается на два подвопроса: (1) Существует ли что-то в реальности, что делает их истинными или ложными? (2) Если подобные делатели-истинности (*truth-makers*) для данного класса утверждений существуют, то располагаются ли они исключительно в объективной реальности, или (по крайней мере, отчасти) и в субъективной тоже?

Отвечая на первый вопрос, мы стоим перед выбором: следует нам отнести пробабилистическое утверждение « $P(A) = n$ » к тому же семантическому классу, что и обычное « A » (если $n > 1/2$) или « $\neg A$ » (если $n < 1/2$)? Или разумнее считать « $P(A) = n$ » вообще не утверждением и не отрицанием, а выражением некоей незаконченной мысли, которая не имеет самостоятельного значения и представляет собой лишь неуверенность, эпистемическое колебание индивида S между мыслями об A и $\neg A$? В первом случае будем говорить об онтологической трактовке вероятностей, во втором – об эпистемологической.

Как уже было сказано, сводить вероятностные оценки целиком к несовершенствам нашего познания было бы большой ошибкой. В то же время, ясно, что какое-то отношение к механизмам познания они имеют. Но какое именно? Если мы принимаем онтологическую трактовку, то вопрос (2) ставит нас перед совершенно новым выбором. Теперь мы выбираем уже не между онтологией и эпистемологией, а между онтологиями разного типа². Один тип представляют «публичные» онтологии, в которых специфическая позиция, занимаемая индивидом S , не существенна при иденти-

² Это деление идейно восходит к отстаиваемому Хинтиккой делению двух способов идентификации объектов (см. [6]).

фикации события A , другой тип – субъект-центрированные³, «перспективные» онтологии, в которых событие A само по себе не всегда может быть идентифицировано, так что специфическая позиция индивида S иногда оказывается решающей (поскольку она задает уникальную эпистемическую перспективу, не доступную автоматически всем остальным индивидам).

Должны ли мы для вероятностных утверждений строить онтологию «от 3-го лица», или «от 1-го лица»? Говоря словами Томаса Нагеля, признаем ли мы наличие в мире специфических реалий наподобие «каково-это-быть- S » (см. [1]), без которых наше знание о мире в принципе может оказаться неполным? В первом случае мы получим «объект-центрированную» онтологию для вероятностей, во втором – «субъект-центрированную».

2. Байесовская вероятность

Принято различать байесовскую концепцию вероятности и частотную. Обе они разделяют одну и ту же математическую аксиоматику, но сильно расходятся в её интерпретации. Байесовская концепция, на наш взгляд, тесно связана с «перспективной» онтологией, частотная – с «публичной». Согласно Байесу пробабилистическая оценка представляет собой не меру возможности наступления некоторого события, а индикатор уровня доверия субъекта к высказыванию об этом событии. Это доверие, поскольку оно имеет четкую меру и основано на рациональном математическом ожидании, следует отличать от психологического ожидания. Психологическое ожидание есть то, что субъект в себе обнаруживает эмпирически, путем интроспекции, в то время как математическое ожидание конструируется теоретически, с опорой на некоторые априорные оценки.

Безусловно, байесовский подход предполагает весьма идеализированное представление о рациональном субъекте: S трактуется как существо, способное (1) строгим образом фиксировать степень своей уверенности в различных высказываниях и (2) столь же строго производить количественную ревизию этой уверенности при получении релевантных эмпирических свидетельств. Процесс такой ревизии опирается на понятие условной вероятности $P(A|B)$, то есть вероятности наступления события A при истинности гипотезы

³ Такие онтологические конструкции чаще всего описываются в терминах «центрированных» возможных миров – см. [7].

тезы В. Собственно, теорема Байеса⁴ утверждает, что апостериорная вероятность события А при условии В (где В – релевантное эмпирическое свидетельство) равняется произведению апостериорной вероятности В при условии А на априорную вероятность А, делённую на априорную вероятность В:

$$P(A|B) = P(B|A) \cdot P(A) / P(B)$$

Другими словами, соотношение апостериорных вероятностей $P(A|B)$ и $P(B|A)$ равно соотношению априорных вероятностей $P(A)$ и $P(B)$ соответственно. Апостериорная вероятность выступает в байесовском подходе как элемент внутренней логики субъекта: она отвечает за внутреннюю убедительность выдвигаемых им гипотез и рациональный характер их пересмотра в связи с вновь полученными данными.

Важный момент, связанный с формулой Байеса, заключается в следующем: при неправильной оценке второго соотношения (что случается весьма часто, поскольку люди не склонны придавать большое значение количественному различию между малой и *очень малой* вероятностью, оценивая их примерно одинаково) субъект легко может неверно оценить и первое соотношение. В результате логически неправильное вероятностное рассуждение может показаться ему интуитивно приемлемым, и наоборот. Это обстоятельство ещё раз подчеркивает, что пробабалистическим высказываниям сопоставляются не психологические, а математические ожидания.

Таким образом, байесовский подход можно описать тремя принципами⁵:

- 1) Субъекты разделяют определенную степень уверенности (*credence*) в тех или иных пропозициях.
- 2) Эта степень уверенности называется вероятностью, и она может быть условной или безусловной. Другими словами, для каждого субъекта в определенный момент времени имеется функция вероятностной оценки P , которая сопоставляет (по крайней мере некоторым) пропозициям H их

⁴ Эта теорема была обнаружена в архиве английского теолога и математика Томаса Байеса (1702–1761) после его смерти и опубликована его другом, философом Ричардом Прайсом. В XX веке заложенная в ней идея персоналистской интерпретации понятия вероятности была развита в работах Рамсея (1931), Сэвиджа (1954), де Финетти (1972).

⁵ См. [3, 589].

В.В. Горбатов. Что приснилось «спящей красавице»?

безусловную вероятность $P(H)$, и (по крайней мере некоторым) парам пропозиций условную вероятность $P(H|E)$.

- 3) Когда субъект оценивает априорную безусловную вероятность гипотезы H как $P(H)$, а априорную условную вероятность H при обнаружении свидетельства E как $P(H|E)$, то получив (релевантное) свидетельство E , он должен пересмотреть вероятность H таким образом, что для него апостериорная вероятность $P'(H)$ будет равна $P(H|E)$.

Долгое время считалось, что одна из главных проблем байесовского подхода состоит в том, что для его практического применения требуется гигантское количество вычислений. Действительно, формула полной вероятности при таком подходе должна включать в себя настолько большое число переменных, связанных с историей эмпирических наблюдений за системой, что её использование вряд ли представляется возможным. Да и сама величина априорной (безусловной) вероятности на практике чаще всего оказывается неопределимой, поскольку субъект склонен игнорировать даже значительные погрешности при её оценке. Однако развитие математического аппарата, равно как и значительный процесс компьютерных вычислений позволяют на сегодняшний день активно применять байесовские оценки в научной статистике – медицинской, экономической, социологической и пр.

Основная трудность байесианства, как оказалось, заключается совсем в другом: можем ли мы непротиворечивым образом совместить наши представления о (i) рациональном субъекте, производящем вероятностные оценки и делающем на их основе сознательный выбор между имеющимися в его распоряжении альтернативами, (ii) логическом устройстве реальности, предоставляющей в распоряжение субъекту эти альтернативы, и (iii) способе их сопряжения в единую систему.

3. Проблема «спящей красавицы»

Насколько серьезна описанная трудность, можно оценить на примере мысленного эксперимента, известного как «проблема спящей красавицы»⁶. Суть этого примера, автором которого считается Адам Элга [4], описывается следующими условиями:

⁶ Elga A. Self-Locating Belief and the Sleeping Beauty Problem // Analysis. 2000. Vol. 60(2). P. 143–147.

- Спящая Красавица (СК) участвует в эксперименте, в ходе которого её погружают в сон, а затем будят (один раз или два раза). Допустим, СК заснула в воскресенье⁷.
- СК знает, что количество её пробуждений зависит от подбрасывания (симметричной) монеты: если выпадет орёл, то её разбудят один раз (в понедельник); если решка, то её ожидают два пробуждения (в понедельник и вторник).
- СК знает, что в случае выпадения решки после первого пробуждения ей введут медикамент, который заставит её забыть про то, что она просыпалась (имеется в виду данное конкретное пробуждение, причем забывает она не только обстоятельства, но и сам факт пробуждения).
- Эксперимент происходит в закрытой комнате без каких либо источников информации о текущем дне недели. Таким образом, СК знает, что когда она проснется, у нее не будет никакой возможности определить, понедельник это или вторник.
- СК знает принципы байесовской концепции вероятности и полностью разделяет их.
- Рассмотрим момент, когда СК проснулась. Как должна она сразу после пробуждения оценить вероятность (*credence*) выпадения орла?

В сущности, вариантов ответа только два: « $\frac{1}{2}$ » и « $\frac{1}{3}$ ». В неформальных обсуждениях сторонники первого называют себя *halfers*, а сторонники второго обозначаются как *thirders* (от «half» и «third», соответственно). На русский язык это различие чаще всего переводится как различие «двоечников» и «троечников». Какой же взгляд является правильным?

Дело в том, что следующие два рассуждения представляются, на первый взгляд, одинаково убедительными:

– Аргумент «двоечников»

До того, как СК заснула в воскресенье вечером, она априори знала, что вероятность выпадения орла $\frac{1}{2}$ (ведь монета симметричная). Когда она просыпается в понедельник утром, то по сравнению с предыдущим вечером она не получила *никакой новой информации*. Единственное, что она «узнала» – так это то, что она дейст-

⁷ В дальнейшем воскресенье будем нумеровать как 0, понедельник как 1, а вторник – как 2.

вительно проснулась; но ей ведь заранее было известно, что она проснётся в любом случае, независимо от того, как выпадет монета. Если монета выпала решкой, то СК получает медикамент и забывает о первом пробуждении, чтобы во вторник утром проснуться снова. И в этом случае после пробуждения она не получила никакой новой информации. Можно, конечно, указать на то, что она потеряла некоторую информацию – а именно, информацию о своем первом пробуждении. Но по условию, она никак не может узнать, что она эту информацию потеряла, и следовательно, с её субъективной точки зрения никакого изменения информации не произошло. Таким образом, у нее нет никаких оснований пересматривать априорную оценку вероятности выпадения орла, и она должна по-прежнему придерживаться ответа « $\frac{1}{2}$ ».

– Аргумент «троечников»

Представим, что эксперимент провели множество раз. Тогда в долгосрочной перспективе число пробуждений СК при выпадении орла (Н-пробуждений) будет в два раза больше, чем при выпадении решки (Т-пробуждений)⁸. То есть число Н-пробуждений, будет составлять *ровно треть* от общего числа пробуждений. Так как СК не может отличить Н-пробуждения от Т-пробуждений, в момент любого пробуждения ей следует пересмотреть свою оценку вероятности выпадения орла, снизив её с $\frac{1}{2}$ до $\frac{1}{3}$.

Действительно, несложно доказать, что все три вероятности, о которых идет речь в этом примере, попарно равны: сначала доказывается $P(T_1) = P(T_2)$ (рассматривая случай, когда СК узнает, что выпала решка), потом $P(H_1) = P(T_1)$ (рассматривая случай, когда СК узнает, что сегодня понедельник)⁹.

Критикуя данное рассуждение, «двоечники» обычно указывают на то, что нельзя допускать подмены вопроса. На вопрос «Какова вероятность выпадения орла?» существует, считают они, только один правильный ответ: $\frac{1}{2}$. Ответ «троечников», кажется, предполагает совсем другой вопрос: «Какова вероятность пробуждения от выпавшего орла?» – на этот вопрос правильным ответом будет $\frac{1}{3}$. Дело, по всей видимости, обстоит так, что если вопрос задан относительно монетки, то пространство возможных исходов состоит

⁸ От англ. «Heads» – орел и «Tails» – решка.

⁹ Более подробные формальные выкладки см. в [4].

лишь из двух альтернатив – Н (орёл) и Т (решка), но при этом альтернатива Т *считается* два раза (как если бы само слово «решка» произнесли дважды – от этого слово не перестаёт быть одним и тем же). А если вопрос задан относительно позиции определенного пробуждения СК в ряду всех её пробуждений, то здесь уже совсем другое пространство возможных исходов: Н1 (орёл-понедельник), Т1 (решка-понедельник) и Т2 (решка-вторник).

Однако «троечники» справедливо отклоняют эту критику как эпистемологически нерелевантную. С их точки зрения, «двоечники» пытаются провести онтологическое различие там, где СК его не может провести, и тем самым искажают смысл проблемы. Даже если в онтологии «от 3-го лица» мы имеем дело с двумя различными событиями (А) «выпадение орла [вероятность чего я оцениваю, проснувшись]» и (В) «мое пробуждение в результате выпадения орла», то с точки зрения красавицы это *ровно одно и то же* событие, вероятность которого оценивается ею в момент пробуждения. Для нее самой высказывания «[Я проснулась] и СК проснулась и выпал орёл» и «Я проснулась и выпал орёл» абсолютно неразличимы, что априори гарантировано условиями эксперимента. Действительно, попробуйте поставить себя на место СК и спросите, чем отличаются выражения А и В? Очевидно, что для вас они будут не только экстенционально, но и интенционально неразличимы.

Таким образом, базовые интуиции «двоечников» и «троечников» диаметрально противоположны. Если первые пытаются проинвестировать декомпозицию эпистемологических вопросов путем введения онтологических дистинкций, то вторые, напротив, выводят онтологические различия из эпистемологических. На наш взгляд, более продуктивной будет не эскалация спора онтологии и эпистемологии, не противопоставление структур бытия и структур познания, но попытка выработать такую онтологию, которая была бы чувствительна к эпистемологическим различиям и «дружелюбна» по отношению к различным эпистемическим перспективам. Подобная онтология должна быть, по всей видимости, динамической и агент-ориентированной, а не статической и объектно-ориентированной.

4. Что узнала «спящая красавица»?

Привнесем в ситуацию элемент азарта. Допустим, красавице предлагают ставить деньги на выпадение орла, и говорят, что эксперимент со сном может быть повторён достаточно большое коли-

чество раз (при этом не важно, будет ли действительно серия экспериментов долгой, или он произойдет лишь однажды – главное, чтобы СК могла отнестись к этой истории как к игре и попыталась выработать некую рациональную стратегию. По тому, какую стратегию она предпочтет, мы можем судить о выносимых ею вероятностных оценках.

Если верна позиция «двоечников», то стратегия ставить на орла ничем не уступает стратегии ставить на решку – в половине случаев она выигрывает, в половине проигрывает. Кажется, ставить на орла разумно, потому что лучшей стратегией для неё все равно не существует. Но необходимо учесть, что в случае выпадения орла красавицу будят (и принимают ставку) только один раз, а в случае решки – дважды. Выбрав стратегию ставить на орла, она будет выигрывать только один раз из трех, а проигрывать дважды. Значит, лучшей стратегией для неё будет ставить на решку. Таким образом, есть серьезные основания считать, что знание субъекта о соотношении числа пробуждений при различных результатах подбрасывания монетки все же играет роль в оценке вероятности и принятии решений. Разумный субъект непременно осознаёт, что даже если он ставит на орла, на самом деле он ставит на свои пробуждения при условии выпадения орла, и значит, должен учитывать сведения, имеющиеся у него на этот счет.

По сути, речь в этом примере должна идти не столько о событиях с монеткой, сколько о событиях с СК и её знаниями. Видимо, в момент пробуждения красавицы ничего не происходит с монеткой и её знаниями о монетке (*de re*), и даже ничего не происходит с её знаниями о высказываниях про монетку (*de dicto*). Но кое-что всё же происходит с её знанием о себе (*de se*): она узнает, что *она* проснулась *здесь-и-сейчас*.

Как отмечает сам Адам Элга, проблема «спящей красавицы» ставит перед нами трудный вопрос о релевантности информации, оформленной в различных модальностях (*de re* / *de dicto* / *de se*) при вынесении вероятностных оценок. Одной из главных его целей было продемонстрировать, что вопреки мнению Льюиса¹⁰, в данном примере информация, имеющаяся у субъекта относительно его собственного положения в пространстве и времени (а точнее,

¹⁰ Lewis D. A Subjectivist's Guide to Objective Chance, in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*. Vol. II. Berkeley: University of California Press, 1980. P. 263–294.

неполнота этой информации) может быть релевантной для пересмотра вероятностной оценки при переходе от стадии предсказания (воскресенье) к стадии вынесения оценки (понедельник/вторник).

5. Двумерный подход

Другими словами, в какой-то момент ситуация предполагает переход от обычных пропозиций к центрированным, и этот переход не является само собой разумеющимся. Чтобы понять двойственную природу знания, полученного красавицей в момент первого пробуждения, необходимо различать не просто экстенционалы и интенционалы (как было показано выше, привычной карнаповской семантической пары понятий тут недостаточно), но интенционалы разного уровня.

Так, например, Дэвид Чалмерс в рамках своей двумерной семантики (см. напр., [2]) выделяет два измерения значения – первое измерение соответствует референту (экстенционалу) рассматриваемого выражения, второе измерение отражает то, как в произвольно выбранном возможном мире референция рассматриваемого выражения зависит от различных внешних факторов (например, от самой структуры этого мира, от характера языковой игры, от обстоятельств произнесения и т.п.). Использование второго измерения призвано сделать значение выражений более гибким и чувствительным к контексту: в зависимости от обстоятельств употребления (прагматических контекстов, эпистемических перспектив, сценариев) одному и тому же выражению могут приписываться различные интенционалы.

Применительно к повествовательным предложениям вторичным интенционалом принято считать множество возможных миров, в которых данное предложение истинно в рамках того эпистемического сценария, который задан действительным миром (подразумевается, что при выборе другого эпистемического сценария истинностная оценка предложения могла бы серьезно отличаться). Первичный же интенционал, по Чалмерсу, можно представить как функцию из множества эпистемических сценариев в множество первичных интенционалов. Именно первичные интенционалы способны быть онтологическим коррелятом того «центрированного знания», которое в примере со «спящей красавицей» служит опорой для позиции «троечников». Спор между «троечниками» и «двоечниками» в таком случае оказывается, по сути, спором о том,

В.В. Горбатов. Что приснилось «спящей красавице»?

как соотносятся первичные и вторичные интенционалы. Почему «центрированное знание» о собственном пробуждении оказывается релевантным вопросу об оценке вероятности определенной нецентрированной пропозиции о монетке? Вероятность какой пропозиции оценивает СК во время пробуждения – ещё нецентрированной, или уже центрированной?

* * *

Возвращаясь к двум трудным вопросам о природе пробабилистических утверждений, которые были поставлены в начале статьи, мы готовы предложить следующее решение: объектом вероятностной оценки должны служить все-таки не обычные (вторичные) интенционалы – онтологические конструкции, скроенные из (нецентрированных) возможных миров и не чувствительные к *de se* установкам, но первичные интенционалы, сделанные из «центрированных миров» и обладающие ярко выраженной эпистемологической окраской.

Литература

1. *Нагель Т.* Каково быть летучей мышью // Хофштадтер Д., Деннет Д. Глаз разума. Фантазии и размышления о самосознании и о душе / Пер. с англ. Эскиной М.А. Самара: Бахрах-М, 2003. Гл. 24.
2. *Chalmers D.* The Foundations of Two-Dimensional Semantics // Garcia-Carpintero M., Macià J. Two-Dimensional Semantics. Oxford: Clarendon Press, 2006. P. 55–140.
3. *Chalmers D.* Frege's Puzzle and the Objects of Credence // *Mind*. 2011. Vol. 120 (479). P. 587–635.
4. *Elga A.* Self-Locating Belief and the Sleeping Beauty Problem // *Analysis*. 2000. 60(2). P. 143–147.
5. *van Fraassen B.* Fine-grained opinion, conditional probability, and the logic of belief. *Journal of Philosophical Logic*. 1995. Vol. 24. P. 349–77.
6. *Hintikka J., Simons J.* Systems of Visual Identification in Neuroscience: Lessons from Epistemic Logic // *Philosophy of Science* 70: 89–104, 2003.
7. *Lewis D.* A Subjectivist's Guide to Objective Chance, in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*. Vol. II. Berkeley: University of California Press, 1980. P. 263–294.
8. *Lewis D.* Sleeping Beauty: Reply to Elga // *Analysis*. 2001. Vol. 61. P. 171–176.

Сведения об авторе: Горбатов Виктор Викторович, старший преподаватель факультета философии НИУ ВШЭ.