

МОДЕЛИ ИЗМЕНЕНИЯ УБЕЖДЕНИЙ: НЕКОТОРЫЕ ОТКРЫТЫЕ ПРОБЛЕМЫ

Н.П. Козаченко

Семнадцатый век был обилён на новации в различных областях. Именно в то время, названное впоследствии Новым, в философии был осуществлён так называемый *эпистемический поворот*. Рассуждения о Боге и пределах мира сменяются поисками границ познания и трактовками о сущности знания. Окружающий человека мир бесконечен и разнообразен и усилий человечества, направленных на его исследование, затрачено было и будет немало. В то же время, человеческая жизнь конечна и стоит ли расходовать её на попытки достижения необозримого? Может быть, проще и целесообразнее начать именно с процесса познания — узнать, каковы познавательные способности человека, каковы возможности и пределы его чувств и разума? Таким образом, онтологические основания знания и разумения оттесняются на второй план.

К примеру, Дэвид Юм, «вместо того, чтобы время от времени занимать пограничные замки и деревни», предлагает «брать приступом столицу или центр . . . — саму человеческую природу», предполагая, что «став, наконец, господами последней, мы сможем надеяться на лёгкую победу и надо всем остальным» [5, с. 56]. Эта прекрасная и далекоидущая идея дала толчок развитию эпистемологии на многие годы вперед, предварительно разделив лагерь исследователей на «скептиков» и «критиков». Поиски скептиков, возглавленных Декартом, были ознаменованы перманентным сомнением во всем, кроме собственно существования, поиском незыблемых столпов нашего знания, к которым может быть вновь применён научный метод. Юм же считал недостаточным сам метод преодоления скептицизма в целом, ибо, сомневаясь во

всем, мы также сомневаемся и в сомнении.

Более плодотворным оказался критический путь, развитый в работах Канта. Исходя из самого факта наличия знания, основное внимание следует уделить именно отделению собственно знания от «сновидения ума» и «сновидения чувств» [2, с. 267], учитывая то обстоятельство, что «человеку свойственно ошибаться». Свести ошибку к такому минимуму, которым можно было бы и пренебречь — вот в чем состоит задача философии. Вследствие этой тенденции возникли и развились разнообразные теории, направленные на прояснение аспектов знания. Разработан ряд концепций истинности, имеются различные подходы к разъяснению структуры знания. Достаточно показательным является то обстоятельство, что знанию, как одному из наиболее общих понятий, довольно тяжело дать исчерпывающее определение. Фактически к каждому определению может быть выдвинут контрпример, хотя и не снижающий его познавательной ценности.

В 1951 году Фон Вригтом были анонсированы такие направления как «эпистемическая логика» и «доксастическая логика», что зафиксировало качественно новый уровень в исследовании механизмов знания. В работе «Очерки о модальной логике» [11] Фон Вригт и разводит эти два термина, назвав эпистемической логику знаний, а доксастической, в свою очередь, — логику убеждений. В самом деле, исследуя природу человеческого знания, поневоле вспоминаем о человеческом праве на ошибку. Может потому и не найдено пока точное определение знания, что человеку свойственно заблуждаться. И при этом он прекрасно оперирует своими «сновиденьями ума и чувств», все далее продвигаясь в познании и себя и окружающего мира. Отличение знания от полагания / мнения / сновидения / заблуждения волновало исследователей еще до нашей эры. Эта тема, так или иначе, затрагивалась в работах практически всех философов, исследующих феномен знания. Также следует упомянуть присутствие феномена веры в познавательной деятельности субъекта. Это — общепризнанный факт, но истоки веры в знании требуют глубоких исследований. Поэтому остановимся именно на рассмотрении убеждения, как более широкого понятия, чем знание. Если мы будем понимать знание только в строгом смысле, то это, считает В.Н. Костюк, «в общем случае препятствует рассмотрению возможности развития знания, перехода от менее полного к более полному знанию, игнорирует элемент гипотетичности в (научном) знании» [3, с. 131]. Основательно такой подход был разработан Хинтиккой в работе «Знания и убеждения» 1962 года [9].

Таким образом, в дальнейшем, употребляя термин «знание», мы

будем подразумевать знание именно в широком смысле, как убеждение (ср. [4, с. 6]). Термин «убеждение» более пригоден для построения систем логического анализа знаний. И если семнадцатый век задавался, главным образом, вопросом Канта «Что я могу знать?», то современная эпистемология направляет свои усилия на раскрытие именно механизмов приобретения и развития знания. Аппарат эпистемической логики позволяет ставить и успешно решать задачи выявления формальных (логических) свойств операторов знания и убеждения (а значит и соответствующих понятий), формулировки аксиом, выражающих эти свойства и установления взаимосвязи между данными операторами и понятиями. Одним из направлений эпистемической логики есть исследование динамики развития знания. Для анализа такого рода аспекта знания применяют метод построения моделей. На сегодня существует целый ряд таких моделей-теорий, позволяющих в той или иной мере описывать процессы изменения знаний. Одна из них, получившая в литературе название *AGM*, основана на работах Алчуррона, Герденфорса и Макинсона и посвящена теории критического пересмотра убеждений (*belief revision*) [6]. Именно на примере этой модели мы рассмотрим некоторые основные проблемы построения эффективных представлений изменения знания, поставленные известным шведским исследователем Свенном-Ове Хансоном в его работе 2002 года «Десять философских проблем пересмотра убеждений» [8].

1 Идеализация

Несомненно, что при построении любой модели, будь то модель аэроплана или развития знаний, применяются различного рода упрощения. Именно упрощения позволяют отделить существенные аспекты исследуемого явления и определить более узкие рамки области изучения. Уже рассматривая сам термин «убеждение» мы применяем упрощение, нивелируя любой личностный аспект приобретения и содержания убеждения. При исследовании развития знания на второй план отходит отношение субъекта носителя знания к его содержанию. Более того, рассматривая множество убеждений, — набор убеждений субъекта, к нему выдвигают требования рациональности, которые ясны и интуитивно понятны, но в то же время существенно ограничивают возможность смоделировать действительный процесс развития знаний человека. Одним из таких требований есть обязательное принятие всех следствий своих убеждений, что в реальной действительности часто не

представляется возможным. В таких случаях и применяется идеализация как выполнение упрощения ради ясности. Если в этом смысле говорят, что модель множества убеждений — идеал, имеют в виду нечто, существующее только как умственная концепция. Идеализация в этом смысле предполагает преднамеренное упрощение некоторой сложной ситуации (концепции и т.д.) с целью достижения, по крайней мере, частичного понимания. Это может повлечь искажение оригинала, или же это просто подразумевает опускание некоторых компонентов ради лучшего сосредоточения на оставшихся.

Но в то же время Хансон обращает наше внимание и на другое возможное истолкование идеализации и идеала. Ведь идеализацией мы можем назвать интерес или наличие слишком высокого мнения относительно чего-либо. А формальные модели не всегда могут выразить нечто, что является совершенным или превосходным в своём роде.

Хансон указывает на то, что формальные модели пересмотра убеждения, такие как *AGM*, были получены с помощью обоих типов идеализаций

1. Упрощенное идеализирование, в результате которого получаем модели, которые не учитывают многие из сложностей реальной жизни;
2. Совершенствованное идеализирование, которое производит образцы, хотя и удовлетворяющие стандартам рациональности, но слишком высокие для реально существующих (доксистических) субъектов.

Уровень разработанности модели, равно как и уровень идеализации — важная характеристика системы пересмотра убеждения. Достаточно часто мы сталкиваемся с рассмотрением рационального идеала субъекта теории изменения убеждений, предполагая наличие у него неограниченной познавательной способности. Это не проблема для формальных моделей, в которых субъект должен обработать бесконечные объекты (вроде бесконечных наборов предложений). Это наиболее общий подход, и математически наименее сложный. Но можем ли мы с помощью этого подхода рационально изучать реальный мир?

Альтернативное представление состоит в том, что идеальные субъекты теорий изменения убеждений обладают ограниченной познавательной способностью, которую они рационально используют. С этой точки зрения, важными достоинствами формальных моделей являются конечность и исчисляемость. Это — аргумент в пользу использова-

ние моделей с конечным представлением состояний убеждений, какими есть конечные базы убеждений. Однако конечность является очень слабым ограничением, требующим от нас реалистического подхода к познавательной способности, но не охватывающим закономерностей динамики убеждений в общем случае.

Таким образом Хансон формулирует первый из своих вопросов:

Существует ли способ представления более строгих познавательных ограничений, чем конечность множества убеждений?

2 Объектный язык

Следующей проблемой, рассматриваемой Хансоном, есть содержание самого набора убеждений. И здесь мы так же сталкиваемся с идеализацией. Убеждения субъекта должны быть представлены наиболее удобным способом. В теории изменения убеждений должен быть объект, который собственно и предполагается изменять. Таким объектом и есть множество убеждений познающего субъекта. Вопрос заключается в том, каким образом должны быть представлены убеждения, входящие в этот набор и существуют ли способы его расширения. Рассматривая само множество убеждений, мы искусственно изолируем убеждения от других элементов разума, — эмоций, предпочтений и т.п. Такого рода модель требует точного отражения состояния убеждений субъекта познания. Большинство доступных моделей изменения убеждений сентенциональны: убеждения представлены исключительно предложениями (высказываниями). Ясно, что это также является идеализацией. Фактические убеждения не обязательно имеют структуру предложений обычного языка. Однако, хотя высказывания не охватывают все аспекты убеждений, они обеспечивают наилучшее и наиболее доступное представление убеждений в целом.

Далее встает вопрос о возможности расширения объектного языка. В большинстве моделей изменения убеждений предложения, которые представляют убеждения, принято считать элементами простого функционально-истинностного логического языка — высказываниями. Соответственно, операции, выполняемые над множеством убеждений, тоже легко представляемы в рамках логики. Но что происходит при добавлении в объектный язык неассерторических высказываний. Включение модальных или условных предложений в язык приводит к интересным результатам, — указывает Хансон, — но одновременно и затрудняет дальнейшую формализацию. В самом деле, содержание объе-

ктного языка — это, прежде всего убеждения, «приспособленные» для выполнения стандартных операций изменения, описанных в модели. Хансон обращает наше внимание на то, что если мы включаем модальные предложения в набор убеждений, стандартные условия изменения могут не выполняться. Это видно на примере постулата включения для сокращения: $(K \div \alpha \subseteq K)$. Пусть высказывание считается возможным, если некто не верит в его отрицание, хотя можно упомянуть и другой смысл возможности, согласно которому некто верит во что-то, для чего имеется достоверное подтверждение (или ряд подтверждений). Тогда получим что предложение «Возможно, что α не есть причиной» может быть элементом $K \div \alpha$ но не K . По-видимому, здесь мы «теряем логику», обогащая язык. То есть дальнейшая формализация предполагает изменение или уточнение операций, определяемых на данной модели. Эта проблема хорошо известна, однако до сих пор является открытой. Исходя из описанного, Хансон формулирует следующий вопрос:

Как могут быть представлены модальные и условные предложения?

3 Степени убежденности

Понятие убеждения может быть истолковано в довольно жестком смысле: либо вы верите во что-то, либо вы не верите в это. Таким образом, получим четкое разделение убеждений и не-убеждений. Субъект — носитель множества убеждений сам отделяет их, веря либо не веря. Все что входит в его множество убеждений и есть убеждения, то что не вошло в это множество, отбрасывается как не стоящее внимания (недостоверное). Согласно такому подходу, модели изменения убеждений дихотомичны: они делят предложения языка на две категории: предложения, которые представляют убеждения и предложения которые не представляют собой убеждений и соответственно не входят в множество убеждений.

Но среди убеждений субъект также естественным образом проводит различия. Чему-то он верит в большей степени, о чем-то только слышал и при поступлении новой информации может с легкостью изменить именно это, менее достоверное убеждение. Такие градации убежденности представляют немалый интерес, как в познавательном смысле, так и в плане формальной реализации. Их можно представить как существование различных степеней убежденности: субъект может быть уверен в своих убеждениях в различной степени.

Хансон в своей работе рассматривает два понятия степени убежденности. Одно из них представляет собой статичную концепция степени доверия. В этом смысле, степень убежденности субъекта в достоверности предложения тем выше, чем более он уверен в убеждении. Другое понятие — это динамическая концепция степени сопротивляемости к изменению. В этом смысле степень убежденности тем выше, чем труднее изменить это убеждение.

Стандартная теория вероятности не проводит никакого различия между этими двумя понятиями, поскольку формально это сделать довольно затруднительно. Однако, некоторые намеки на их различие могут быть найдены в литературе по теории изменения. Айзек Леви проводит тщательное различие между уверенностью и неизменяемостью (*incorrigibility*) [10]. Это различие не может быть осуществлено в рамках теории вероятности, так как там рассматриваются высказывания с определённой степенью вероятности, фактически аналогично степеням уверенности. Уверенность субъекта может быть обоснована различными способами, будь то логическое доказательство, правильность которого определяется правилами вывода и истинностью, опять же, убеждений, использованных при этом, либо, к примеру, способом получения этого убеждения.

Сопротивляемость к изменению определяется иными параметрами, сходными с характеристиками бинарного оператора укорененности, предложенного Герденфорсом.

Предложенное Хансоном различие, между уверенностью и сопротивлением к изменению, могут быть представлены как обобщение различия Леви. Кроме того, понятие эпистемического укоренения может быть истолковано как попытка интерпретировать сопротивление к изменению. Ведь именно эпистемическая укорененность предложения в эпистемическом состоянии субъекта определяет судьбу убеждения при пересмотре. Определение оператора укорененности явно указывает, какое убеждение более устойчиво к любого рода изменениям множества убеждений.

В свете этого, вышеупомянутое положение о том, что стандартные модели изменения убеждения дихотомичны, Хансон интерпретирует более точным образом. Модели дихотомичны относительно степеней убежденности, но не относительно степеней сопротивляемости изменению.

Дихотомическая картина далека от бесспорной. Согласно Байесовому идеалу рациональности, рациональный субъект не должен иметь слишком много полных убеждений. Только логически истинные пре-

дложения могут иметь вероятность, равную единице. Конечная система убеждений — сложная сеть связанных утверждений различной степени вероятности. В действительности же такая система убеждения была бы неуправляемой для человека. Хансон указывает что наши познавательные ограничения настолько серьезны, что переход от высокой вероятности до полной уверенности неизбежно должен сделать нас способными к осуществлению заключений и принятию решений на этой основе. Другими словами, мы обращаемся с убеждениями, как будто стопроцентно уверены в них, хотя это далеко не так. Это приравнение к полному убеждению, или «фиксация убеждения», помогает нам достичь познавательного управляемого представления мира.

Распространенность этого процесса фиксации является одной из причин того, что дихотомические модели убеждений представляют некоторые особенности докстического поведения субъекта (особенно ценного с точки зрения логики) более реалистично, чем вероятностные модели. Очевидно, что есть другие особенности, которые могут быть более реалистично представлены именно в вероятностных моделях. Обратим внимание, однако, что этот аргумент о преимуществе дихотомических моделей относится лишь к субъекту с ограниченной познавательной способностью. Использование дихотомических моделей в описании субъекта с неограниченными познавательными способностями, оказывается достаточно узким и поэтому не слишком оправданным.

В результате Хансон подразделяет модели изменения знаний на два типа:

1. Вероятностные модели, в которых не проводится никакого различия между уверенностью и сопротивляемостью к изменению; оба типа градаций представлены одной и той же функцией вероятности;
2. Связанные модели, основанные на укоренённости, в которых степени сопротивляемости к изменению представлены отношением укоренённости, принимающие во внимание отсутствие различия между степенью уверенности, с которой убеждения поддерживаются.

По мнению Хансона следует точно выяснить фактические отношения между двумя понятиями степеней убежденности. Возможно ли сведение их в одну концепцию? Именно в этом и заключается следующий вопрос Хансона.

Каковы формальные и неформальные отношения между двумя понятиями степени убежденности — обоснованности и сопротивляемости к изменениям?

4 Уязвимость и оправдание

Далее речь пойдет именно о дихотомических моделях теории изменения убеждений. В таких моделях, отношение между множеством убеждений и предложениями, в которые верят, может быть выражено с помощью функции поддержки s . Эта функция отбирает предложения, которые поддерживаются состоянием убеждений. Пусть K — состояние убеждений. Тогда $s(K)$ — набор всех предложений, которые являются убеждениями в K .

В теории изменения убеждений логические следствия из убеждений обычно принимаются как самостоятельные убеждения, то есть считается, что $s(K)$ замкнуто относительно логического следования $s(K) = Cn(s(K))$, где Cn — оператор логического замыкания (присоединения следствий). Замкнутость множества убеждений познающего субъекта — не слишком реалистичное предположение, но оно чрезвычайно полезно для получения управляемой формальной структуры. Интересный аргумент в его пользу был выдвинут Айзеком Леви, согласно ему, результат $s(K)$ должен интерпретироваться как состоящий из таких предложений, в которые кто-то обязуется верить, но не обязательно верит в них фактически.

Самое простое и наиболее очевидное представление состояний убеждения должно идентифицировать каждое состояние убеждений с его соответствующим набором убеждений, так, чтобы $s(K) = K$. Тогда операции изменения могут выполняться на наборе убеждений. Точнее — на некоторых основных убеждениях, а не на тех, которые из них могут быть получены. Этот случай аналогичен модели *AGM*: набор убеждений считается равным состоянию убеждений. Однако, это описание *AGM* по мнению Хансона, вводит в заблуждение. В модели теории изменения убеждений статической конструкции набора убеждений недостаточно для описания состояния убеждений. Мы должны знать не только убеждения, подтверждаемые сейчас, но также и судьбу этих убеждений после выполнения некоторых операций по их изменению.

В литературе исследованы два основных способа представления этой динамической информации. Один из них должен осуществлять прямое представление сопротивляемости изменению, или наоборот, уя-

звимости элементов первоначального набора убеждения, указывая, каким образом оставляются различные убеждения. Это — доминирующий подход к пересмотру убеждений в стиле *AGM*. Идея состоит в выборе и сохранении менее уязвимых предложений. Уязвимость не зависит от специфики действий, которые будут выполнены (и, следовательно, от входного предложения).

Предположим, что есть два оператора \otimes_1 и \otimes_2 , и два предложения a и b , которые являются кандидатами на удаление для каждого оператора. Тогда, согласно подходу уязвимости, более уязвимо чем b относительно оператора \otimes_1 , если и только если выполняется также относительно оператора \otimes_2 . (Для конкретности, мы можем представить \otimes_1 как удаление $a \wedge b$, и \otimes_2 — как удаление $a \wedge b \wedge c$).

Другой тип динамической информации касается оправдательной структуры набора убеждений. Некоторые убеждения не имеют независимого статуса, а поддерживаются только с помощью других убеждений. Когда оправдывающее убеждение удаляется, оправдываемое убеждение возможно также должно быть удалено. Есть два главных способа выразить информацию об оправдательной структуре. Самый простой из них использует базы убеждений. База убеждений это набор предложений, который (кроме граничного случая) не замкнут относительно логического следствия. Его элементы представляют собой убеждения, которые поддерживаются независимо от любого другого убеждения или множества убеждений и могут даже противоречить друг другу. Логически замкнутая база убеждений представляет собой множество всех убеждений. Элементы набора убеждения, которые не присутствовали в базе убеждений, а «просто получены», не имеют независимого положения. Изменения, выполненные над изначальными предложениями базы убеждений отличаются от изменений «полученных» убеждений.

Хансон подчеркивает, что базы убеждений обеспечивают только очень грубое приближение к оправдательной структуре. Они охватывают лишь дедуктивные оправдания, которые включают только часть фактических оправдательных отношений.

Более реалистическая оправдательная структура содержится в способе, который можно назвать представлением хранения следа. Здесь, к каждому предложению некоторым образом добавлен (присоединен) список из его оправданий или происхождения. Хранящие след представления кажутся особо подходящими для реализации фактической оправдательной структуры, но набор посылок (как например, база убеждений) также может быть использован в качестве материала для по-

строения потенциальных оправдательных структур. Можно выбрать что-то и более подходящее: выражать ли оправдательную структуру, набор возможных происхождений или набор возможных посылок.

Отношение между уязвимостью и оправдательной структурой остается открытой проблемой. Хансон указывает, что остается неясным, на каком уровне степени убежденности — концептуальном или техническом — оправдательная структура может быть выражена в терминах уязвимости или наоборот.

Какое отношение между уязвимостью/сопротивлением и оправдательной структурой?

5 Категориальное соответствие

В дальнейшем рассмотрении формальных представлений состояний убеждений следует обратить внимание на возможность выражения изменений состояния в формальной структуре. Здесь также могут быть рассмотрены различные варианты представления. Одним из таких представлений могут быть временно-индексированные модели (дискретные или непрерывные), зависящие от переменной времени. Изменение состояния мира, положения вещей или состояния убеждений может быть представлено как функция переменной времени. Эта структура может также быть истолкована индетерминистски, с учетом связки функций, типично структурируемых как выполняющих переход по принципу дерева. Этот подход был полезен во многих областях исследования, но редко используется в изучении изменения убеждений. Ведь за одно изменение переменной времени система может претерпеть целый ряд изменений состояния убеждений, вызванных определенным перечнем внешних влияний и комплексных воздействий.

Вместо временного представления, большинство моделей изменения убеждений построены как модели, непосредственно обрабатывающие входные изменения (*input-assimilation*). В таких моделях, изменяемый объект зависит от входа, и изменяется в результате обработки этой информации. Никакое явное представление времени не включено. Вместо этого присутствует характерный математический элемент — функция, которая каждой паре состояния убеждения и поступившего входа, приписывает новое состояние. В хорошо построенной обрабатывающей вход модели результат изменения должен иметь тот же формат, что и до изменения. Этот принцип назвали принципом категориального соответствия [7]. Если мы изменяли базу убеждений,

то в результате должны тоже получить базу убеждений. Точно так же, если первоначальное состояние убеждений — это набор убеждений, объединенный с информацией об уязвимости его элементов, новое состояние убеждений после изменения должно содержать те же самые два элемента, а не превратиться, например, в набор убеждений без сопровождающей информации об уязвимости.

Категориальное соответствие оказывается простым и интуитивно понятным, но труднодостижимым критерием. Например, оригинальная модель АГМ не удовлетворяет этому принципу. Практически рассматриваются множества убеждений вместо пар, состоящих из набора убеждений и функции выбора по укорененности. И в результате изменения системы изменяется и множество значений функции укорененности. Таким образом Хансон обращает внимание на серьёзные трудности в достижении категориального соответствия. И этот вопрос также немаловажен именно в применении к модели АГМ.

Как лучше изменить модель АГМ, чтобы достигнуть категориального соответствия?

* * *

Всего Хансон предлагает к обсуждению 10 открытых вопросов, но рассмотренные в предложенной статье проблемы уже открывают обширное поле для исследований. Дальнейшие вопросы построения моделей изменения знаний, затронутые Хансоном, предполагают более конкретное рассмотрение их на прикладных системах, поэтому будут освещены в последующих публикациях.

6 Литература

- [1] *Декарт Р.* Сочинения: в 2-х т. — Т. 1. — М.: Мысль, 1989.
- [2] *Кант И.* Сочинения в 6-ти т. — Т. 3. — М.: Мысль, 1964.
- [3] *Костюк В.Н.* Элементы модальной логики. — К.: Наукова думка, 1978.
- [4] *Шрамко Я.В.* Знания и убеждения: их развитие и критический пересмотр // *Философия науки.* — 2005. — № 1. — С. 3-19.
- [5] *Юм Д.* Сочинения: в 2-х т. — Т. 1. — М.: Мысль, 1965.

- [6] *Alchourron C.*, *Gärdenfors P.*, *Makinson D.* On the logic of theory change: partial meet contraction and revision functions // *Journal of Symbolic Logic.* — 1985ю — Vol. 50. — P. 510-530.
- [7] *Gärdenfors P.* and *Rott H.* Belief revision // *Gabbay D.* et al. (eds.) *Handbook of Logic in AI and Logic Programming.* — Oxford University Press, 1995. — Vol. 4. — P. 35-132.
- [8] *Hansson S.O.* Ten philosophical problems in belief revision // *Journal of Logic and Computation.* — 2003. — Vol. 13. — P. 37-49.
- [9] *Hintikka J.* *Knowledge and Belief.* — Ithaca: Cornell University Press, 1962.
- [10] *Levi I.* *The Fixation of Belief and Its Undoing.* — Cambridge: Cambridge University Press, 1991.
- [11] *von Wright G.H.* *An Essay in Modal Logic.* — North-Holland, 1953.